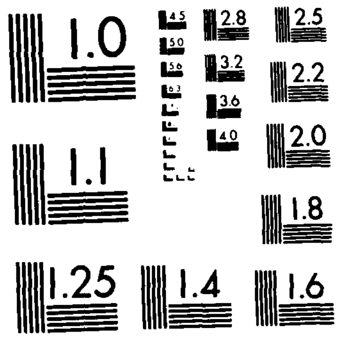
AD-A172 738 THINKING ABOUT NON-LINEAR SMOOTHERS(U) PRINCETON UNIV NJ DEPT OF STATISTICS J W TUKEY MAY 86 TR-291-SER-2 ARO-23360.4-MA DAAL03-86-K-0073 1/2

UNCLASSIFIED F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A172 738

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

DTIC ELECTE OCT 0 8 1986 S D

D

1a. REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE:

3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S):

5. MONITORING ORGANIZATION REPORT NUMBER(S): ARO 23360.4-MA

6a. NAME OF PERFORMING ORGANIZATION: Princeton University

6b. OFFICE SYMBOL *(If applicable)*:

7a. NAME OF MONITORING ORGANIZATION: U. S. Army Research Office

6c. ADDRESS *(City, State, and ZIP Code)*: Princeton, NJ 08540

7b. ADDRESS *(City, State, and ZIP Code)*: P. O. Box 12211, Research Triangle Park, NC 27709-2211

8a. NAME OF FUNDING / SPONSORING ORGANIZATION: U. S. Army Research Office

8b. OFFICE SYMBOL *(If applicable)*:

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DAAL03-86-K-0073

8c. ADDRESS *(City, State, and ZIP Code)*: P. O. Box 12211, Research Triangle Park, NC 27709-2211

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*: Thinking About Non-Linear Smoothers

12. PERSONAL AUTHOR(S):

13a. TYPE OF REPORT: Technical

13b. TIME COVERED: FROM ______ TO ______

14. DATE OF REPORT *(Year, Month, Day)*: May 1986

15. PAGE COUNT: 116

16. SUPPLEMENTARY NOTATION: The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*: Nonlinear Smoothers; Nonlinearity; Smoothing; Linear Smoothing

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

DTIC FILE COPY

Any kind of smoother is not easy to grapple with, either to understand or to choose, but non-linear smoothers - - often the smoothers to be preferred - - are harder to grasp than the simpler, linear ones. The purpose of this account is to give its readers some background with which to think about non-linear smoothers, particularly resistant ones. It does not attempt the task - - probably today quite unfeasible - - of providing a comprehensive guide to which smoother to use where and when.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL:

22b. TELEPHONE *(Include Area Code)*:

22c. OFFICE SYMBOL:

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

ARO 23360.4-

Thinking about non-linear smoothers

by

John W. Tukey

Princeton University
Fine Hall
Washington Road
Princeton, NJ 08544

Technical Report No. 291, (Series 2)
Department of Statistics
Princeton University

May 1986

Prepared in part in connection with research at Princeton University sponsored by the Army Research Office (Durham), DAAL03-86-K-0073.

Thinking about non-linear smoothers

John W. Tukey

Princeton University
Fine Hall
Washington Road
Princeton, New Jersey 08544

TABLE OF CONTENTS

Thinking about non-linear smoothers

John W. Tukey

Technical Report No. 291, (Series 2)
Department of Statistics
Princeton University
Princeton, New Jersey 08544

****** 1. Introduction ******

Any kind of smoother is not easy to grapple with, either to understand or to choose, but non-linear smoothers - - often the smoothers to be preferred - - are harder to grasp than the simpler, linear ones. The purpose of this account is to give its readers some background with which to think about non-linear smoothers, particularly resistant ones. It does not attempt the task - - probably today quite unfeasible - - of providing a comprehensive guide to which smoother to use where and when.

* non-linearity? *

The word "non-linear" does not look too different from the word "linear", but similarity of appearance covers up a tremendous difference in scope. Think of the earliest days of the ancient Greeks, when their ships never went outside the Mediterranian Sea - - and the then difference between "Mediterranian" and "non-Mediterranian". As Western history evolved "non-Mediterranian" grew to include the Bay of Biscay, the East Coast of Africa, the Atlantic, Indian and Pacific Oceans and distinctive land areas on many continents. More recently areas on the moon, and limited aspects of the surface of a number of planets have to be included. What "non-Mediterranian" covers is now much more diverse than what "Mediterranian" ever covered, and the relative diversity is still growing. The relation of "non-linear" to "linear" - - in any field, not just in smoothing - - is like that of "non-

Prepared in connection with research at Princeton University sponsored by the Army Research Office (Durham) through DAAL03-86-K-0073.

Mediterranian" to "Mediterranian". So we ought to expect the discovery and exploration of one interesting area after another - - some which are quite similar to "linear" and some of which are quite different. We will need new tools - - in the Mediterranian, the Greeks had little need for either ice axes or parachutes - - and new ways of looking at the phenomena we uncover.

It is not easy to remember that the non-linear might prove to be infinitely more diverse than the linear, but we ought to try.

* smoothing and smoothers *

The processes of smoothing - - and the algorithms that carry them out - - surely have purposes, but it is often not easy to be explicit what these purposes are. (We will return shortly to some of them.) And it is quite clear that

a) there are qualitatively different purposes,

b) they often have to be compromised, AND

c) quantitatively different compromises of the same purposes are often needed.

As a result, even *linear* smoothing involves a broad repertory of detailed processes and algorithms - - and is not at all easy to think about. Making choices among linear smoothers is not easy; the writer knows of no book that explains "how to choose" in a really helpful manner. (Often, no linear smoother is able to do what is needed.)

With both "smoothing" and "non-linear" in such difficult hard-to-handle states, is it any surprise that thinking about their combination "non-linear smoothers" is not easy? And will not be made easy by reading this paper? Or by reading any book that can be conceived today?

* some purposes *

There are a diversity of purposes for which smoothing seems appropriate. Some of them can be identified without too much trouble, including:

d) taking the "sharp corners" off data to be plotted, so that the viewer's eye-and-brain (often abbreviated "eye") can see appropriate general aspects of the data's behavior better (otherwise isolated points, for instance, often seize more attention than they deserve),

e) ridding the data of much of the irrelevant variation that contributes to each of its numbers, without disturbing too seriously the slower changes that reflect the changing underlying causes that are, in those particular instances our real concern,

f) preparing the data for further processing, especially for further processing that - - like the eye - - would be oversensitive to irregularities.

g) separating, and setting aside, more rapid changes from less rapid ones, at least to whatever degree is possible.

These purposes may sound rather similar, but close scrutiny - - especially of the smoothers to which they lead - - will show not only their distinctness, but a great diversity of need within each of them. We will try, in this paper, to help with thinking about purposes and about the relation of choices to purposes, but all of us need to admit that there is no substitute for practice - - and especially for practice that leads, many times over, to *comparison of the effects* of different examples of such choices on either real or simulated data - - better on both.

Further purposes that may not, at least at first glance, seem like smoothing are:

h) preserving the breaks or sharp corners that might prove important, while eliminating the little wiggles that are likely to distract the eye, AND

i) catering to parsimony by replacing heavily smoothed results by closed form functions expressed by simple formulas.

But these really do belong to the same broad class of purposes.

The relation of smoothing to forecasting is thought to be simple and close by some, but less so by others.

* modes of description *

How do we want to describe

smoothers = processes of smoothing

in a way or ways that will be most helpful? The answer here is equally not straightforward. To explain why, we will gain by listing the more obvious modes in which we often need to describe a smoother (which we assume has already been given a label):

j) *Algorithms* - - descriptions of the details of the successive steps from input to output,

k) *Strivings* - - what properties/behavior we have tried to build into each of our smoothers, and how vigorously we have pursued them,

m) *Benchmarks* - - how each of our smoothers behaves - - qualitatively and quantitatively - - in a well-chosen set of standard situations,

n) *Properties* - - what we can say, in varying generality, about how each smoother performs - - this may be qualitative or quantitative, and is likely to overlap, to a limited degree, with "Benchmarks".

We are, in most subareas, early in our study of non-linear smoothers. As as consequence, we often have to emphasize algorithms, and perhaps strivings. If we knew more, we would be able to emphasize benchmarks and properties, which would be to our great advantage. Just looking at an algorithm - - even for one experienced in smoother design - - is a poor way - - often a very poor way - - to understand how the smoother in question will perform.

Clearly we - - or someone - - has to know an algorithm, else we or our computers would not be able to apply it. However,inferring very much about behavior directly from the algorithm is not at all easy - - often it is impossible. The algorithm makes the label realizable. Only trial - - perhaps by ourselves on a limited number of examples, but not infrequently, fortunately, by others on more extensive and more diverse examples, is likely to lead to useful insight into its detailed behavior, since few aspects of general behavior have so far proved accessible to mathematical argument, even for some smoothers or some components of them. (Most smoothers that will interest us here are assembled from components.)

* plan *

The body of this account, which now follows, tries to develop two frameworks; one for kinds of description, and one for the presently most attractive classes of smoothers, in the hope that the two will help each of us in thinking about non-linear smoothers and non-linear smoothing. Both explicit discussion and examples will be confined to one-dimensional smoothing, but we need to notice that some of the more valuable applications are to two-dimensional data -- usually to images.

Detailed descriptions and characteristics of individual smoothers are at most mentioned as examples. (At some later time, some extension, perhaps an appendix to this account, might arise to present such information.)

* scope *

While, as just noted, something is known about smoothing for values scattered in the plane, etc., we will here only be concerned with smoothing of finite sequences, where the data consists of a finite set of numbers indexed by integers or by more or less regularly spaced numbers (ties among the index values, however, not excluded).

There is, in principle, an important distinction between equi-spaced and non-equi-spaced sequences. There are times when we do recognize this distinction. But the behavior of many of the methods that we discuss does not seem responsive to this distinction. As a result, we have often to recommend treating non-equally-spaced sequences in the same way we would recommend if they were equally spaced. This is particularly true with median-based smoothers.

PART I. SOME KINDS OF BEHAVIOR

****** 2. Problems and strivings ******

Strivings, here as elsewhere, arise as we struggle with problems. So we ought to begin with some of the clearly recognizable problems.

* a short problem list *

It is now time, therefore, to identify some of the most prominent technical problems, with the intention of shortly discussing each in turn:

a) *erosion* - - the tendency of smoothers, especially naive ones, to "wear down the peaks and fill in the valleys".

b) *tenting* - - the tendency of linear smoothers to respond to a single, exotically high value by constructing a "tent" below it, and, by symmetry, to respond to a single, exotically low value by constructing an inverted tent above it.

c) *diversity* - - the fact that a particular property of a smoother may be an advantage in some situations, but a disadvantage in others

d) *balance* - - the need, in choosing a smoother, to balance incommensurables - - as when greater smoothness of result requires the smoothed values to be not as close to the originally given values ("balance" seems more elegant than "compromise", but the idea is the same).

* erosion *

The existence of erosion causes many smooths to be shrunk toward a common value, global or sectional. To correct this, we need to begin by comparing, in some way, the smooth with the data. One simple and useful way is to introduce the *rough*, according to the identity

$$\text{data} \equiv \text{smooth} + \text{rough}$$

and to seek evidence for needed modification of the smooth from the behavior of the rough.

If we find systematic behavior in the rough, it is natural to want to transfer that systematic behavior from rough to smooth. Often, the simplest way to do this is to smooth the rough, and then start from the two identities

$$\text{data} \equiv \text{smooth} + \text{rough}$$

$$\text{rough} \equiv (\text{smooth of rough}) + (\text{rough of rough})$$

and to substitute the second in the first, inserting appropriate brackets, to reach

$$\text{data} \equiv [\ \text{smooth} + (\ \text{smooth of rough}\ )] + [(\text{rough of rough})]$$

It is now natural to take

$$\text{new smooth} = \text{smooth} + (\text{smooth of rough})$$

$$\text{new rough} = \text{rough of rough}$$

and to describe the process as *reroughing*. (If the second smoother is the same as the first, we alternatively refer to the process as *twicing*.)

Many ways of dealing with erosion that were initially described in other ways can be put into the form of reroughing. Any kind of correction that depends *only* on the values of the rough - - anything which does not look at the smooth - - is a process that accepts a sequence - - the rough - - and produces a sequence consisting of: the values to be taken out of the rough for insertion in the smooth. This process, since it generates a smoother sequence from an input sequence (here the first rough) can be regarded as a smoother. Its application can thus be considered reroughing.

If we are to seek more general ways of dealing with erosion, then, we must look at the smooth as well as the rough. This means that we need to try to distinguish peaks, that will be cut down, from valleys, that will be filled up - - and to distinguish both from upward or downward inclines. One simple approach, not supposed to be perfect or even highly effective, would be to look at a second difference of the smooth, spread out over a moderate range of the index.

If we adopt

$$\begin{aligned} y_i &= \text{data} \\ z_i &= \text{smooth} \\ r_i &= \text{rough} \end{aligned}$$

where, of course,

$$y_i \equiv z_i + r_i$$

we could look at the values of such expressions as

$$H_3(i) = \frac{+z_{i-3} - 2z_i + z_{i+3}}{6}$$

$$H_4(i) = \frac{+z_{i-4} - 2z_i + z_{i+4}}{8}$$

or their analogs - - or some combination of these - - embedding them in some so-far unspecified algorithm.

While these might be useful in building, probably after combination with appropriate values of the rough, an effective erosion compensator for a linear smooth, we are likely to need a modified approach when dealing with non-linear smoothers.

For some of the simpler non-linear smoothers, we might consider

$$K_3(i) = \text{median}\{-(y_i - y_{i-3}), 0, y_{i+3} - y_i\}$$

$$K_4(i) = \text{median}\{-(y_i - y_{i-4}), 0, y_{i+4} - y_i\}$$

and so on, which only respond quite near either the top of a peak or the bottom of a valley. Little, if anything, seems to have been done about using such erodibility indicators, either alone or in conjunction with the values of the rough.

It is far from clear, however, whether there are practical circumstances where the influence of reroughing away from peaks and valleys is unfortunate. Thus we do not really understand where, if anywhere, we would want such modified processes of transfer from rough to smooth.

* tenting *

If we take the simple sequence with a single exotic value, 144, 96, 132, 144, 108, 84, 60, 72, 48, *1200*, 48, 24, 36, 50, 48, 84, 96, 132, 120, 144 and smooth by running means of 3

$$z_i = \frac{y_{i-1} + y_i + y_{i+1}}{3}$$

we get the sequence ?, 124, 124, 128, 112, 84, 60, *440*, *432*, *424*, 36, 40, 48, 64, 76, 104, 116, 132,? which shows the rather square "tent" . . .small, 440, 432, 424, small, . . . in place of the single exotic value . . .small, 1200, small, ... Further linear smoothing will spread the tent out, probably slanting its edges somewhat, but the total size of the tent will continue to resemble the roughly 1150 of the original single exotic value's deviation from the general run of its neighbors. No linear smooth will get us away from this effect.

The simplest way around tenting is to replace linear combinations by more robust summaries. The simplest of these are running medians, as when

$$z_i = \mathit{median}\,\{y_{i-1}, y_i, y_{i+1}\} \qquad (\text{"3"})$$

$$z_i = \mathit{median}\,\{y_{i-1}, y_{i-1}, y_i, y_{i+1}, y_{i+2}\} \qquad (\text{"5"})$$

or, when we are willing for the smoothed values to come half-way between adjacent data values, as in

$$z_{i+1/2} = \mathit{median}\,\{y_i, y_{i+1}\} \qquad (\text{"2"})$$

$$z_{i+1/2} = \mathit{median}\,\{y_{i-1}, y_i, y_{i+1}, y_{i+2}\} \qquad (\text{"4"})$$

A single isolated exotic value will be almost forgotten by "3", "5" or "4", but not by

"2".

We can, of course, make use of other robust summaries, such as biweights, or hubers. These are only likely to be chosen when we want to smooth more vigorously, and are looking at 8 or more values of y at a time.

There are also important methods involving the robust fitting of straight lines, etc.

* diversity *

Some data sequences behave as if they had a break at some intermediate position in the sequence. The apparent break may be a change in level - - or a change in slope - - or something more complicated. The prototypic example of a change in level, uncomplicated by any irregularity, is something like

- - - 0, 0, 0, 0, 0, 0, 100, 100, 100, 100, 100, 100, - - -

Such smoother components as "3" or "5" will leave this break untouched (and the whole sequence unaffected). Others, like "2" repeated, will do their best to put in a smooth transition between 0 and 100. We *cannot say generally which of these behaviors we prefer*. For some kinds of data and some purposes we clearly prefer to have the break preserved - - for others we prefer a smooth transition.

The same is true of breaks in slope - - we will discuss an example in section 9 where it seems very natural to preserve breaks in slope, and, conversely there are many instances where this is not the case.

The question of breaks is only one of a number of questions where the direction of preference depends upon kind of data and kind of purpose. The main lesson to be learned from these issues of diversity is that we dare not look for a single chosen smoother, to be recommended for use in any arbitrary situation. We must offer the user a decent palette of smoothers - - and guidance in choosing among them. This means, most importantly for our present concern, that the user has to expect to do some *thinking* about alternative smoothers - - and that the user ought to expect to try more than one smoother on the same data whenever the details of the outcome are important.

* further diversity *

After the qualitative choices that we have just been discussing come a variety of quantitative choices - - shall we use a smoother based upon "3" or one based upon "5"? - - shall we rerough only once, or do it again? These are often more difficult than the qualitative choices. All that we know how to do so far is to *try* to "include enough small-scale diversity in our palettes, without being excessive". Just how we ought to set about making up such palettes is *not* something that has been adequately considered.

* balance or compromise *

In the present case, our problem is complicated by incommensurability of what we are striving for - - the largest-scale-instance of which is

reaching a smooth result, AND

keeping close to the original data

These are aims that obviously tend to pull our choice in almost opposite directions. What is hard to face - - and a rock on which organized compromise can easily founder - - is the apparent absence of any natural way to write down

a measure of lack of smoothness, AND

a measure of deviation from the original data

that are either in, or convertible so as to be in, comparable units.

In classical robustness as applied to location, we have had to face a similar, much easier problem. When we are happy to work with performance under each of 2 or 3 situations, which we are happy to compromise, we face the fact that, for instance:

variance (or MSE) for a standard Gaussian, AND

variance (or MSE) for the standard slash

are not directly comparable. (Here the standard slash is the distribution of a unit Gaussian divided by an independent unit rectangular [0,1].) In the first instance, we can deal with this by asking what is the best - - the smallest variance or MSE - - that we can do for Gauss alone or for slash alone, and then going over to

% excess variance

(excess over the minimum we know how to attain) both for Gauss and for slash (or for each of the few situations that we consider).

Having done this, a first natural thing to do seems to become minimaxers, to seek a compromise that minimizes the maximum % excess variance (for two alternatives, this means equating the two % excess variance). While it has not yet become customary to go further than to seek a single compromise, it may throw light on our present, more general problem if we try to take another step.

As a tentative proposition, in the case of only two alternatives, let us think about proceeding as follows:

If the minimax % excess variance is E, identifying the symmetric compromise, let us consider two *satellite* compromises (satellite in the spectroscopic sense), in each of which one % excess is allowed to grow to $E\sqrt{2}$, while the other is made as small as possible. (If we wish to go further, going to a % excess of 2E for one alternative is conveniently called a *dim satellite*.

This satellite construction can be carried out for either a one-parameter family of estimates or some larger class.

For the n =20 Gauss-slash compromise, this produces, for the one-step biweight family -- using the graphs in Bell and Morgenthaler, 1981 --

label	*tuning constant*	*excess* at Gauss	*excess* at slash
satellite	5.5	22%	7.6%
symmetric	6.5	15%	15%
satellite	7.8	8.7%	22%
(dim satellite)	9	3.1%	31%

and for estimates bioptimal among all equivariant estimates

label	*shadow ratio*	*excess* at Gauss	*excess* at slash
satellite	2.1	6%	2.5%
symmetric	1.29	4.3%	4.3%
satellite	.67	3.2%	6%

where the "shadow ratio" defines the linear combination of the two % excess variances whose optimization gives the indicated estimates.

This whole approach is heavily undergirded by two facts

- the two criteria to be compromised have been made satisfactorily comparable by changing from raw variance to % excess, AND
- the % excesses involved are all small (in our examples no more than 15% for the symmetric compromises).

When we try to use explicit compromises in the smoothing situation, it is not clear that either of the analogous facts holds for any reasonable way of re-expressing our two measures of dissatisfaction.

It is possible, though it is not clear whether the details can be carried out, that we can come to a comparable situation in the following indirect way:

- Let us define a smallest tolerable amount of smoothing, and measure deviation of smooth from the given data, as a % increase over this smallest amount (a robust measure of deviation size, perhaps like s_{bi}^2, will be required).

- Let us define a largest tolerable amount of smoothing, and measure lack of smoothness as a % increase of roughness over that corresponding to this "heavy" smooth.

- Then let us play the "satellite, symmetric, satellite" game.

Clearly no one knows whether or not this is a reasonable approach (without regard to whether its result would be successful). It requires four difficult choices; two of criteria and two of degree: criteria of lack of smoothness and of poorness of fit, and greatest (because deviations from what was observed are otherwise unacceptable) and least (because of lack of smoothness is otherwise unacceptable) degrees of smoothing. Moreover, the compromised % excesses probably cannot be allowed to be too large.

We have suggested an approach for two reasons:

- it seems an effective way to make the difficulty of the problem clear, AND

- it may encourage the suggestion of other approaches.

* non-singleness *

An essential in current treatments of robustness, and in the approach to formal compromise in smoothing just considered, is the focusing on single aspects - - in the examples above on a pair of single aspects.

In the robustness-of-location instance, focusability was not obviously guaranteed. We accepted the % excess variance measure, itself based on a variance measure, because the shapes of the distributions of estimation errors of different high-performance estimates are surprisingly similar. This is a bonus, whose existence we have recognized as a consequence of much tedious experimental sampling and of careful analysis of the results of such sampling; a bonus whose very existence seems still to be beyond easy explanation. Even in that single instance, we could hardly have counted on focusability in advance of experimental sampling - -

even though we were dealing with distributions of error for single numbers.

When we come to deal with the smoothing instance, our situation is much worse. Our concern is not just with a single output value, nor is it even with each of the output values singly. There are many important aspects of quality of the output that are much more holistic, either sectionally or globally. We have to look seriously at $z_k, z_{k+1}, \ldots, z_{k+m}$ as a whole, not just as a collection of separate values. Indeed, we have to do this more importantly for the z's than for the y's.

This is a type of criterion-invention problem with which we have inadequate experience. So we need to push on and get some. This means not just writing down criteria - - much of that has been done to little avail. It means coming much more closely to grips, initially in verbal and vague terms, with what lack of smoothness ought to mean to us and why. (We do not attempt this here.)

****** 3. Near linearity ******

* IS-boxes *

We use "box" to refer to any well-defined process with one or more inputs and an output.

A one-input "box" that is both *superposable*, namely satisfies

output from a+b = (output from a) + (output from b)

and *invariant* under changes in time origin

output from (a shifted in time by h) = (output from a, shifted in time by h)

is conveniently called an IS-box, I for Invariant and S for Superposable. The notion of an IS box formalizes what is often called *linearity*. Thus IS boxes make up the Mediterranian from which we start.

If we are dealing with a sufficiently nearly linear processes, or, more generally, with polynomial processes, we may find it appropriate to describe important aspects of non-linear processes, including some non-linear smoothers, through simple (or simple-seeming) modifications of the definition of IS-boxes.

* quadratic and bilinear boxes *

Following Tukey 1984 (Volume 1) pp. 584ff we shall use [ ] to denote the output of a *(homogeneous)* quadratic box, where the input is given in the brackets. The simple identity

$$[a+b] + [a-b] = 2[a] + 2[b]$$

for all inputs "a" and "b" and their sums and differences is a simple and effective way to define what is quadratic without bothering about details. (This approach to polynomiality traces at least to the classic papers of Mazur and Orlicz (1935) on polynomial operations).

From the identity it is easy to show (see *ibid* pp. 584-585) that

$$[0] = 0$$

and

$$[ka] = k^2[a]$$

for all rational k. Now only a touch of continuity is needed to give this relation for all real k.

If we define $\langle\ ,\ \rangle$ by

$$2\langle u, v\rangle = [u+v] - [u] - [v] \qquad a \qquad (*)$$

it is easy to show *(ibid* pp. 585-588) that

$$\langle a+b, c+d\rangle = \langle a, c\rangle + \langle b, c\rangle + \langle a, d\rangle + \langle d, d\rangle$$

so that $\langle\ \ \rangle$ is linear in each of its inputs and is thus conveniently called bilinear.

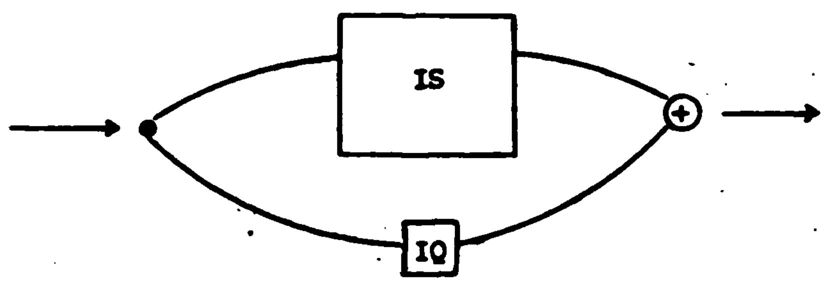
* IQ-, ISS-boxes *

If we are dealing with more general boxes that are also time-origin-shift invariant, we use "IQ-box" for a single-input box that is quadratic in the sense just described and "ISS" for a two-input box that is bilinear (that is, superposable in each input separately. A simple consequence of what we have indicated above (at (*)) is that:

- given a few copies of an IQ-box, we can make an ISS-box
- given a few copies of an ISS-box, we can make an IQ-box,
- if we follow one these constructions with the other, in either order, we return to an equivalent of the box with which we started.

* linear-PLUS-quadratic boxes *

The gentle approach to non-linearity is to consider boxes that are inhomogeneous quadratic in the sense that their output can be realized as the sum of the outputs of IS and IQ boxes sharing an input.
Schematically, we could write



This is a natural analog of the beginning of a simple power-series expansion. It is easy to understand in frequency terms, as we will see in the next section. There are kinds of non-linearity for which it is a useful beginning.

* IH-boxes - - proportionality *

The statistician - - and, more generally, the data smoother - - is likely to be much more drastic, when he or she considers being non-linear. Think of perhaps the simplest of the non-linear smoothers, namely

running medians of 3

where

$$z_t = \text{median}\{y_{t-1}, y_t, y_{t+1}\}$$

So far as we know, there is no useful polynomial representation - - surely there is no linear-PLUS-quadratic representation - - for this smoother. It is almost utterly non-polynomial.

It does satisfy a condition of homothety (proportionality), namely

output from (k times a) = k times (output from a)

(We probably also want good response to an additive constant, which it has.)

This shows easily that it can't be linear-PLUS-quadratic since any linear piece will satisfy this condition, but no quadratic piece can (they all require k^2 on the right, not k).

When it is convenient to have a notation for boxes that

- are time-origin-shift Invariant, AND
- satisfy the Homothety condition

we will call them IH-boxes. Clearly every IS-box is an IH-box, but not *vice versa*. Clearly the only box that is both IQ and IH is the null box (all of whose outputs are null).

* IP-boxes - - polynomiality *

We could extend the ideas back of quadratic boxes, both homogeneous and inhomogeneous, to more general polynomial boxes. (Orlicz and Mazur have the appropriate identities.) We might use IP- box for any (inhomogeneous) polynomial box. And we would find that the only IP-boxes that are also IH-boxes are the IS-boxes. For references to polynomial boxes in general see page 306 of Brillinger 1970.

In a data-smoothing world where IH-boxes are the rule, focussing our attention on polynomial boxes - - or on more general initial segments of power-series-like representations - - seems doomed to failure. The kinds of non-linearity we want to use are too drastic for such approaches.

*** WS-, WX-, and WP- boxes - - except at the ends ***

Our discussion of "nice" boxes always involved time-origin-shift invariance, involved "shifting an input by h". If this has no other effect than to time-shift the output, presumably this can be done as many times as we wish, something which implies unrestricted (and hence infinite) extent in time for both inputs and outputs. Since we never seem to have inputs of wholly unrestricted length, something has gone awry here. What should be our stance?

Think about something rather simple, say smoothing by running medians of 5

$$z_i = median\{y_{i-2}, y_{i-1}, y_i, y_{i+1}, y_{i+2}\}$$

which, as it stands, is *not defined* when i corresponds to one of the first two or last two values of an input.

We have a choice

- to let outputs be shorter than inputs, OR
- to define graceful degradations of our smoothers near the ends of the input.

Only if we have very long inputs does the first alternative have a reasonable chance of being acceptable. As we shall see, most non-linear smoothers concatenate individual smoothing components. When this occurs, the shortening from the overall process is the sum of the shortenings from the individual components, and may thus be quite large.

So only the choice of some graceful degradation remains. If i goes from 1 to n, for instance, we may start and stop a running median of 5 with shorter running medians

$$\begin{aligned} z_1 &= \text{median}\{y_i\} = y_1 \\ z_2 &= \text{median}\{y_i, y_2, y_3\} \\ z_{n-1} &= \text{median}\{y_{n-2}, y_{n-1}, y_n\} \\ z_n &= \text{median}\{y_n\} = y_n \end{aligned}$$

In addition to such a simple sort of graceful degradation, we may well need some form of further fixup, one that operates close to the ends, such as "the end value rule" (see EDA, Tukey 1977, Chapter 7). (We may be able to use preliminary extrapolation as a route to graceful degradation, but I know of no examples.)

When we want to be careful, we replace

$I =_{df}$ time-origin-shift-invariant

by

$W =_{df}$ time-origin-shift invariant EXCEPT near the ends of the input or output, where the smoother, or more general box, is modified in a planned way.

Superposition, homothety or polynomiality can still be required for inputs of fixed length.

Accordingly, ideal IS-boxes need to be replaced by real WS-boxes, ideal IH-boxes by real WH-boxes, and ideal IP-boxes by real WP-boxes. And ideal ISS-boxes become real WSS-boxes.

This sort of care in labeling represents a care in thought that is always appropriate, and most often necessary.

****** 4. Angular frequencies ******

If we have equally-spaced data $\{y_t\}$, as we have just seen the range of t will always be finite - - and this finiteness will usually matter. This is at least as true in connection with analysis into sinusoids and cosinusoids like

$$C \cos(\omega t + \phi)$$

as any careful discussion of spectrum analysis shows us. As a result (angular) frequency analysis is unlikely to be really helpful in studying the smoothing of short inputs.

With this caution, we shall turn to how such frequency analysis can illuminate the smoothing of "long" inputs, inputs where we are *not* concerned with behavior *near* the *ends* of either input or output.

* transfer functions *

If

$$y_t = C \cos(\omega t + \phi)$$

for some C, w, and ϕ, and if $\{y_t\}$ were to be the input to some IS-box, then the output has to be of the form

$$z_t = D \cos(\omega t + \psi)$$

for the same ω. In more specific words, all an IS-box can do to a single cosinusoid is

- to change its size by a factor D/C, AND
- to change its phase by addition of $\psi - \phi$, WHERE
- these changes do NOT depend upon C or ϕ.

(For proofs for various cases, see Tukey 1984, pp. 507 to 509.)

It is convenient to combine these changes into a complex number $L(\omega)$, where

$$L(\omega) = (D / C)e^{i(\psi - \phi)}$$

where D/C and $\psi - \phi$ are, of course, functions of ω. It is usual to call $L(\omega)$ the *transfer function* of the IS-box.

If we have a finite sum

$$\sum_h C_h \cos(\omega_h t + \psi_h)$$

our IS-box would give as output

$$\sum_h D_h \cos(\omega_h t + \psi_h)$$

something we can calculate from the representation of the input and the values of $L(\omega)$ at the ω_h. Since we can represent any finite stretch of input as such a sum of cosinusoids we can find any finite stretch of output given $L(\omega)$ and a finite stretch of input.

In reality, of course, the best we can ask for is a WS-box, but - - *except near the ends of input and output* - - its behavior will be completely described by the corresponding transfer function.

There may be advantages, in studying the behavior of specific WS-boxes, to supplement the transfer function of the corresponding IS-box by some description of near-the-end behavior, but no systematic way of doing this has attracted the writer's attention.

In more illuminating words, transfer functions completely define IS-boxes because an IS-box does NOT ENTANGLE frequencies - - which means that each frequency in the output comes entirely from the same frequency in the input - - while the same is truc of WS-boxes, except near the ends of the input and output.

* blurred transfer function *

The smoothers we discuss here are not likely to be either IS-boxes or WS-boxes, although they may resemble them in some ways. As a consequence, they do entangle frequencies to a degree, and their behavior is more complicated. To move on to the next approximation, let us suppose that

$$y_i = C \cos(\omega t + \phi) + Y_i$$

and that we have fixed upon a procedure, given output $\{z_i\}$ and frequency ω, to write

$$z_i = D^* \cos(\omega t + \psi^*) + Z_i^*$$

where the output corresponding to $\{y_i\}$ - - the same input minus the cosinusoid — takes the form

$$D^{**} \cos(\omega t + \psi^{**}) + Z_i^{**}$$

Thus, adding "C cos $(\omega t + \phi)$" to the input has added to the output an amount, if we write

$$D^* e^{i\psi^*}$$

to mean amplitude D^* at phase ψ^*,

$$D^* e^{i\psi^*} - D^{**} e^{i\psi^{**}}$$

at frequency ω as well as

$$\{Z_i^{**} - Z_i^{**}\}$$

which we think of as being at other frequencies. Accordingly

$$L(\omega) = \frac{D^* e^{i\psi^*} - D^{**} e^{i\psi^{**}}}{C}$$

is the apparent transfer function, which now depends on the $\{y_i\}$.

We no longer have a single valued transfer function. Rather we have a blurred one. If we wished to insert a probability distribution for the "noise" $\{Y_t\}$ we could have a probability distribution for $L(\omega)$ - - probably most accessible by simulation - - and would naturally tend to consider the average and variance of its values at each ω.

Little has yet been done to introduce this degree of realism.

The importance of such ideas today is mainly to ensure that we do not think of any particular non-linear smoother as having an exact transfer function.

* transport functions *

An IQ-box - - a homogeneous quadratic box - - has the following frequency behavior

$$\omega \;\; IN \;\; \rightarrow \;\; 0, 2\omega \;\; OUT$$
$$\omega_1, \omega_2 \;\; IN \;\; \rightarrow \;\; 0, 2\omega_1, 2\omega_2, \omega_1 + \omega_2, \omega_1 - \omega_2 \;\; OUT$$

An ISS-box - - a bilinear two-input box - - has this frequency behavior

$$\omega_1 \;\; IN_1 \, , \;\; \omega_2 \;\; IN_2 \;\; \rightarrow \;\; \omega_1 + \omega_2, \omega_1 - \omega_2 \;\; OUT$$

An IP-box, say inhomogenous of degree 3, with "ω_1, ω_2, ω_3 IN, that is, with input

$$y_t = C_1 \cos(\omega_1 t + \phi_1) + C_2 \cos(\omega_2 t + \phi_2) + C_3 \cos(\omega_3 t + \phi_3)$$

has an output that may, and is likely to, involve the following frequencies

$$\begin{array}{ll} 0 & \\ \omega_1, \omega_2, \omega_3 & \\ 2\omega_1, 2\omega_2, 2\omega_3 & \\ \omega_1+\omega_2, \omega_1-\omega_2, \omega_1+\omega_3, \omega_1-\omega_3, \omega_2+\omega_3, \omega_2-\omega_3 & \\ 3\omega_1, 3\omega_2, 3\omega_3 & \\ 2\omega_i \pm \omega_j & (i, j, \mathit{any\ two\ of}\ 1, 2, 3) \\ \pm(\omega_1 \pm \omega_2 \pm \omega_3) & \end{array}$$

Once we leave the IS-box, IP boxes can be expected to *transport* input at one frequency (or more frequencies) into output at other frequencies.

What about IH-boxes? There seems

- to be no simple argument as to what sort of transfer ought to take place,
- adequate empirical evidence that input at a single frequency is transported mainly to that frequency and its harmonics
- inadequate insight into what happens when pair or triples of frequencies - - or more complicated sequences - - serve as arguments.

We can usefully start to define a *transport function* $M(\omega \rightarrow \omega')$ by input

$$y_i = C \cos(\omega t + \phi)$$

and output

$$z_i = D\cos(\omega' t + \psi) + Z_i$$

where Z_i is intended to be "free of frequency ω". It is then natural to try to put

$$M(\omega \rightarrow \omega') = \frac{D}{C} e^{i(\psi - \frac{\omega'}{\omega}\phi)}$$

and to have to face the fact that, in general, the right-hand side will depend upon ϕ. (The expression in the exponent may make more sense when we realize a time-origin shift of h has these consequences

$$\begin{aligned} \phi &\rightarrow \phi + \omega h \\ \psi &\rightarrow \psi + \omega' h \\ \frac{\omega'}{\omega}\phi &\rightarrow \frac{\omega'}{\omega}\phi + \omega' h \\ \psi - \frac{\omega'}{\omega}\phi &\rightarrow \psi - \frac{\omega'}{\omega}\phi \end{aligned}$$

showing this expression as the simplest one revealing time-origin-shift invariance.

At the very least then, we have to try to understand

$M(\omega \rightarrow \omega')$ as a function of ϕ

- - as something whose image is a loop, small or large - - especially for $\omega' = \omega, 2\omega, 3\omega, \ldots$. Transport functions will not be easy to understand, and only a beginning on this understanding has been made (see Velleman 1975.)

* blurred transport functions *

All the immediately above was for pure single-cosinusoid inputs. If we are to understand smoother performance for real inputs, it is probable that we will have to go to blurred transport functions.

* intermodulation functions *

When we study those human-built analog-signal boxes that come closest to IS behavior - - hifi amplifiers - - we do not study their transport functions - - though for all we know it might be important to do so. Rather we apply

$$y = C_1 \cos(\omega_1 t + \phi) + C_2 \cos(\omega_2 t + \phi_2)$$

often with widely different ω_1 and ω_2 and look at frequencies $\omega_1 - \omega_2$ and $\omega_1 + \omega_2$ - - looking for "intermodulation". This has served us well in studying amplifiers, we do not know whether or not it will serve us well studying smoothers.

* some dangers *

When one has an input that is likely to include occasional exotic values under circumstances where (linear) filtering would have been appropriate if there were no exotic values, we can think about at least three alternative approaches:

- construct a non-linear filter in a rather direct way, and apply it to the input
- use a robust cleaning procedure to remove the exotic values, and then apply a linear filter,
- repeat cleaning and filtering either in order or in some combined way.

The first of these is often dangerously attractive to the beginner. If one dares to forget the transport and intermodulation behaviors of most non-linear smoothers - - or of more general non-linear filters - - the idea of combining, in a single process, the stripping away of the possible effects of exotic values with the desired filtering seems attractive. But doing it is far from easy.

The special case of monochromatic robust smoothing - - of low-pass filtering where the input is a single sinusoid plus noise (possibly stretch-tailed) was fairly successfully handled by Velleman (1975), but we do not even know how his selected smoothers would perform for a combination of two cosinusoids plus noise.

* a warning example *

Let us look at a fairly simple example. Let our non-linear smoother be running medians of 5

$$z_t = median\{y_{t-2}, y_{t-1}, y_{t-1}, y_{t+1}, y_{t+2}\}$$

and suppose our input is

$$y_t = 100 \sin \frac{2\pi t}{5} + D \cos \frac{2\pi t}{2.2222} + noise$$

where both D and the size of the noise are small.

The values of $100 \sin \frac{2\pi t}{5}$ are 0, 95.11, 58.78, -58.78, -95.11, 0, 95.11, 58.78, -58.78, . . . repeating with period 5. So long as the remainder of $y(t)$ is not too large, say

$$\left| D \cos \frac{2\pi t}{2.2222} + noise \right| < 18$$

the median of any five adjacent y's is that y for which $100 \sin 2\pi t/5 = 0$, that is, for which $t \equiv 0 \pmod 5$.

If t starts at zero, and there is no noise,

$$z_0 = z_1 = z_2 = D \cos 0 = D$$

$$z_3 = z_4 = z_5 = z_6 = z_7 = D \cos \frac{10\pi}{2.2222} = D \cos 4.50\pi = D \cos 0.5\pi$$

$$z_8 = z_9 = z_{10} = z_{11} = z_{12} = D \cos \frac{20\pi}{2.2222} = D \cos 9.00\pi = D \cos \pi$$

$$z_{13} = z_{14} = z_{15} = z_{16} = z_{17} = D \cos \frac{30\pi}{2.2222} = D \cos 13.50\pi = D \cos 1.5\pi$$

$$z_{18} = z_{19} = z_{20} = z_{21} = z_{22} = D \cos \frac{40\pi}{2.2222} = D \cos 18\pi = D \cos 0$$

$$z_{23} = z_{24} = z_{25} = z_{26} = z_{27} = D \cos \frac{50\pi}{2.2222} = D \cos 22.5\pi = D \cos .5\pi$$

$$z_{28} = z_{29} = z_{30} = z_{31} = x_{32} = D \cos \frac{60\pi}{2.2222} = D \cos 27\pi = D \cos \pi$$

etc.

Thus z_t is periodic with period 20, and has a simple wave form. Accordingly a substantial amount of

$$\cos \frac{2\pi}{20} t$$

appears in $\{z_t\}$ - - in fact, this term will be by far the most sizable frequency present.

As well as annihilating the

$$\sin \frac{2\pi t}{5}$$

term, the running medians of 5 have transported energy from the

$$\cos \frac{2\pi t}{2.22222}$$

term, whose frequency of oscillation is 1/2.2222 = .45 cycles/point, to a

$$\cos \frac{2\pi t}{20}$$

term, whose frequency of oscillation is 1/20 = .05 cycles/point. ***Beware of transport and intermodulation.***

*** Mallows' linear closest ***

It is natural to try to study non-linear smoothers by asking which linear smoothers - - which IS-boxes, which transfer functions - - approximate them most closely. If smoothers behaved like IS-boxes with little IQ-boxes in parallel, such an approach might prove very powerful. For smoothers that behave like IH-boxes, however, we must be prepred to be grateful for whatever small gains any such approach can yield. These results have already proved useful in correcting for gentle variations in $L(\omega)$ caused by the use of a non-linear smoother (Schwartzschild, 1979).

And it may be that we can come to understand the essentials of the non-linear behavior of certain boxes, perhaps even certain smoothers, by studying the modified boxes whose final output has been corrected for the linear consequences of their use by applying the inverse of Mallows's closest linear approximation to the initial output.

Colin Mallows (1980) has studied this question. His results are interesting, but

of limited help. He approximates

[non-linear smooth of] (Gaussian signal PLUS white noise)

(where "white noise" means independence from one time point to another) by

[linear smooth of] (same Gaussian signal)

(note the absence of action by the linear smooth on the noise!)

and finds a unique best fitting linear smooth. However, this best-fitting linear smooth depends on *both* which Gaussian signal process *and* which white noise we are presumed to be concerned with. Thus trying to "omit the non-linearities" gives different results for different inputs (to an extent that seems not to have been studied). The "linear closest" is not at all like a transfer function.

These results are limited to the case where signal PLUS noise is white. Again little seems to have been done to study dependence on shape - - and relative size - - of the noise distribution.

Little here seems likely to be easy; probably nothing can be used immediately to provide major increases in our insight.

****** 5. Simple benchmarks ******

Frequency analysis of smoother behavior may eventually be quite powerful, but its use involves complexities and difficulties. Thus, there is an important place for simpler methods, even when these give quite limited information. Of these, the use of benchmarks seems likely to be particularly helpful. We discuss simple, individual-input benchmarks in this section, and more complex, mainly probabilistic benchmarks in the next.

* kinds of simple benchmarks *

The simplest inputs we might use for benchmarks include:

- breaks - - inputs in which one constant value suddenly changes to another
- straight lines - - inputs that decrease or increase linearly
- polynomials (in the time index)
- box cars and towers - - inputs that are zero except for a more or less short stretch where they take a common non-zero value
- binomial bumps - - inputs that are zero except for a more or less short stretch where their values are those of the binomial coefficients $\binom{n}{i}$ for chosen n
- single-color sinusoids - - inputs of the form $C \cos(\omega t + \phi)$
- combinations of the above.

We will now say a few words about each of these in turn.

* breaks *

The desired response of a non-linear smoother to a break is not always the same. Sometimes, especially in image processing, it is of overwhelming importance to *preserve* the breaks. At other times, especially when what underlies the data is reasonably sure to be smooth, it can be of great importance to "smooth over" the breaks - - and thus keep them from distracting the viewer.

Response to breaks is a tool for sorting smoothers appropriate for different uses, rather than a uniformly applicable criterion of quality.

* straight lines *

The input

$$y_t = A + Bt$$

is just about as smooth as an input can be. Thus there is no *need* for a smoother to change such input. Ordinarily, we feel strongly that our smoothers should *preserve* straight lines, turning out an output identical with the input.

* polynomials *

This desire for preservation extends to polynomial of appropriate degree, almost always to quadratics and usually to cubics, sometimes beyond. Polynomials are of interest

- because they are simple to describe and manipulate, AND
- because they imitate, sometimes closely and sometimes not, important aspects of the behavior of either real inputs or of what after being contaminated with noise became the real input.

Thus quadratics simulate individual smooth maxima and smooth minima, sometimes quite well. And cubics can simulate the connection of a smooth maximum and a smooth minimum.

We often would like to have our smoothers preserve polynomials of degree $\leq$ some k, either exactly (an ideal) or approximately (sometimes a reality).

* box cars and towers *

Lewis Carroll may have originated "what I tell you three times is true" (a later science-fiction story describes the effect of including this maxim in a large information system). One of the main purposes of non-linear smoothers is often *not to believe what happens only once*, in other words to pay very little attention to a single wild value.

Some number of adjacent similar values will need to be taken seriously. The proper cutoff - - between what is surely not taken seriously and what will often need to be taken seriously - - will vary from application to application.

A smoother like running medians of 3, which almost neglects a single exotic value, but preserves two equal adjacent exotic values, acts as if "what I tell you twice is true."

A smoother like running medians of 5 acts as if "what I tell you three times is true!" And so on.

Box cars and towers also serve to classify smoothers into groups made up of candidates for different classes of applications.

* binomial bumps *

Besides short constants - - box cars and towers - - it is useful to understand how specific smoothers respond to specific short, but more or less smooth inputs. While broken-line inputs might seem simplest, they do not seem to imitate important aspects of very common inputs. As a result, they do not appear to be a useful benchmark.

The binomial coefficients, which give a tower for n=1, give smoother bumps for larger n (and even approximate a Gaussian density for very large n). The simplest cases are:

0	0	0	0	0	0	1	1	0	0	0	(n=1)
0	0	0	0	0	1	2	1	0	0	0	(n=2)
0	0	0	0	1	3	3	1	0	0	0	(n=3)
0	0	0	1	4	6	4	1	0	0	0	(n=4)
0	0	1	5	10	10	5	1	0	0	0	(n=5)
0	1	6	15	20	15	6	1	0	0	0	(n=6)

(Here the zeroes are part of the input, and continue, in both directions, as far as needed.)

Unless "what I tell you twice is true!" applies we would like our smoother to neglect a binomial bump for n=1. On the other hand, we would like to preserve binomial bumps for large *n*, at least approximately.

The smoothers "3R" and "3R twice" when applied to the binomial bump for n=4, both yield, as outputs,

0 0 0 1 4 4 4 1 0 0 0

and hence as roughs (input MINUS output)

0 0 0 0 0 2 0 0 0 0 0

while "3RH" and "3RH twice" yield,

0 0 .25 1 3.25 4 3.25 1 .25 0 0

and

0 0 .25 1.18 4 4 4 1.18 .25 0 0

respectively, as smooths.

Rather than criteria to be rigidly met, responses to binomial bumps seem to be behavior to be understood, behavior whose understanding often increases our understanding of the overall behavior of the smoother concerned. Again understanding of this behavior may let us sort out smoothers in yet another way.

* single-color sinusoids *

When we want to see behavior on something smooth and moderately simple, but not specifically localized (like a binomial bump), the most natural class of candidates seems to be the single-color sinusoids

$$y_t = \cos(\omega t + \phi)$$

where we often need to look at a fair number of values of ω, starting with rather smooth instances, which arise for small ω.

Since the input is periodic, and the smoother is, probably, W, we are likely to have periodic output (as always, away from the ends of the input and output). Thus we are not likely to need to look at more than 1.5 or 2.0 cycles of output. (Looking at only 1.0 cycles can mislead us.)

With non-linear smoothers, the value of ϕ can matter, although for IH- or WH-smoothers a change of ϕ by π, which takes y_t into $-y_t$ offers no new information.

Thus we may want to look at 2, 3, 4 or possibly more, values of ϕ - - which may well be limited to $[0, \pi]$ - - for a given ω - - in the hope that the corresponding behaviors will not be too different, but not with certainty that this will happen.

Careful thought about how to display the answers may be worthwhile. Generally - - since we are describing smoothers - - we anticipate (near) preservation for small ω and (near) rejection for large ω (in our case of integer t, "large" means ω's approaching π).

* combinations of the above *

There may well be much to learn from combinations of benchmarks of the types just briefly discussed. However, we haven't really started to do this yet.

* closing comment *

He who wishes to understand a specific smoother, or wants to learn to think about smoothers, will do well to calculate what his smoother - - or a few selected smoothers - - do to a variety of simple benchmarks.

****** 6. Distribution-based benchmarks ******

Besides the simple benchmarks, there is a place - - often in combination with simple benchmarks - - for benchmarks which simulate irregular variation, "noise" if you will. Most of these are stochastic - - are thought of as consisting of a population of possibilities and dealt with in terms of a sample - - of some number of realizations drawn at random from the corresponding population.

* Gaussian noises, some white *

At one extreme are the "Gaussian noises" where $y_1, y_2, \ldots, y_n$ have a joint Gaussian distribution, most often a distribution as unaffected by origin-shift as possible, so that $(y_1, y_2, \ldots, y_{n-1})$ has the same distribution as $(y_2, \ldots, y_{n-1}, y_n)$. (This implies that the covariance of y_i with y_j only depends upon $|i-j|$.)

When used in combination with (after superposition on) a simple benchmark, the most frequent case is that of a white Gaussian noise, where all the y_i are independent of one another. this is often a reasonable facsimile of a "nice" background noise.

* stretch-tailed noises - - mostly white *

Background noise need not be nice; in fact a main reason for the existence of non-linear smoothers is the likelihood of exotic values. Two sorts of stretch-tailed noises seem most useful for challenging smoother behavior:

- contaminated Gaussian noise where α% of a broad Gaussian distribution is mixed with (100-α)% of a narrow Gaussian with the same center, AND
- slash noise, which can be generated by dividing a zero-center Gaussian deviate by an independent rectangular deviate (uniformly distributed on [0, A] for some $A > 0$).

Again the "white" case, where y_i is independent of y_j for $i \neq j$, has been used almost exclusively.

These "noises" are also intended to imitate an irregular background. Good smoothers will reduce their effects on the output almost as far as possible.

Good performance against both Gaussian and stretch-tailed noise is almost a *sine qua non* for good robust smoothers.

There are important applications where noises are "bursty" - - where exotic values tend to come in groups of 2, or 3, or more; I have no experience upon which to comment.

* combinations among simple benchmarks *

Here are several opportunities for the future. Velleman's work (1975) focussed on a single cosinusoid plus white noise of different kinds.

PART 2. SOME CLASSES OF SMOOTHERS

****** 7. Median-based components ******

This section introduces, rather briefly, the basic median-based components, and a few modifications. Recall that we met the simplest median-based components in section 2, under "tenting".

* kinds of median *

When we have an odd number of values, say the five values 9, 4, 1, 2, 5, their median is the middle value after sorting in order — (1, 2, 4, 5, 9) — and hence 4 in this example.

When we have an even number of values, say 8, 3, 6, 7, there are *two* middle values, *after* sorting in order, in this example 6 and 7. We call their mean *the* median, the lower one the *lomedian* and the higher one the *himedian*. Thus, for instance

$$\begin{aligned} &\text{med}\{8,3,6,7\} = \frac{1}{2}(6) + \frac{1}{2}(7) = 6.5 \\ &\text{lom}\{8,3,6,7\} = 6 \\ &\text{him}\{8,3,6,7\} = 7 \end{aligned}$$

We extend these rules to negative values directly, so that, for instance

$$\begin{aligned} &\text{med}\{7,-1,-2,-4\} = -1.5 \\ &\text{lom}\{7,-1,-2,-4\} = -2 \\ &\text{him}\{7,-1,-2,4\} = -1 \end{aligned}$$

thus ensuring that for any a and c $\geqslant 0$, and any k $\geqslant 2$

$$\text{med}\{a+cx_1, a+cx_2, \cdots, a+cx_k\} = a+c\cdot\text{med}\{x_1, x_2, \cdots, x_k\}$$

$$\mathrm{lom}\{a+cx_1, a+cx_2, \cdots, a+cx_k\} = a+c\cdot\mathrm{lom}\{x_1,x_2, \cdots, x_k\}$$
$$\mathrm{him}\{a+cx_1, a+cx_2, \cdots, a+cx_k\} = a+c\cdot\mathrm{him}\{x_1,x_2, \cdots, x_k\}$$

(For negative c, the first relation continues while the other two require "lom" on one side and "him" on the other.)

For odd k, the "him" and "lom" of any k values are, of course, the same as their "med".

* warning about "2", "4", ... *

Rather clearly, if we were to plot

$$\frac{1}{2} y_t + \frac{1}{2} y_{t+1}$$

we ought to plot it at $t + \frac{1}{2}$. All running medians (or running means, etc.) of even lengths have this property. It is almost always desirable, therefore, to use such components in pairs, one after the other (still other component smoothers can be put in between, of course) so that our indices move first from integers to half integers, and then back to integers.

* selectors and semiselectors *

Colin Mallows has introduced the term "selector" for a function of k variables whose value is always one of its arguments. Medians for odd k, and all lomedians and himedians, are selectors.

It may prove convenient to define a semiselector as a function of k variables whose value is always EITHER

- one of its arguments OR
- the average of two of its arguments

Clearly all medians are semiselectors.

If we take a selector, and substitute a selector for one or more of its arguments - - where, if we substitute two or more, we may substitute either the same selector or different selectors, but generally with different arguments - - the result is easily seen to be a selector. [A corresponding statement about semiselectors is false.]

* to the death *

Those smoothing components that are selectors are usually also, in a sense which it does not seem helpful to make too precise here, both smoothing and *shrinking*, in the weak senses that their output is both not rougher and not more spread out than their inputs. As selectors, since n y's have at most n different values, their repeated use can produce at most n^n different sequences. So repetition can only lead to eventual constancy or cycling. And cycling will ordinarily be incompatible with "smoothing and shrinking".

Thus, at least for components or subassemblies that are selectors, it makes sense to define "R" as expandable to "repeated to death" or "repeated to no further change" as an instruction to repeat the indicated component or subassembly until no further changes occur. Such a definition is only useful when the needed number of repetitions is small — or possibly moderate. (The frequently observed tendency of continuing change to be concentrated in a few segments, rather than throughout the sequence helps to make a moderate number of repetitions bearable in hand calculation, since we may only need to recompute for a few short stretches.)

The use of R allows simple components to generate much more potent subassemblies. Thus "3" is helpful, though its output has no easily specifiable properties, but "3R" has a simple property - - it leaves alone any output that moves monotonically up — or down — between flats where two or more adjacent values are equal.

* roots *

Whether or not we do "R", we need to have some interest in the classes of sequences left unaffected by a particular smoother. These have been rather felicitously called "roots" of the smoother; for some results see Nodes and Gallagher

(1982) and Huang (1981).

* the $\check{\text{sh}}(\ )$ components *

We have already noticed the importance of a variety of attacks on erosion — and the limited gain to be had by relying on reroughing (esp. twicing) alone. The sequence of components we are about to describe were called into existence by a desire to reduce erosion in the most erosive steps.

* $\check{4}$ *

With a, b, c, d, e five successive values in our sequence, $\check{4}$ is defined as follows (the mark above the digit is intended to be a "hash mark" as in the Czech language):

$$\check{4} \text{ gives, to replace c, } \begin{cases} \text{med}\{b, d\}, & \text{if } (a-b)(d-e) < 0 \\ \text{med}\{a, b, d, e\}, & \text{else} \end{cases}$$

In words, if a,b go up and d, e down or vice versa, so that there seems to be a peak, or a valley, between b and d, we take a median of only the two values b and d, thus going less far down the mountain (or up the valley walls) (than if we had used med{a, b, d, e}. In such situations med{a, b, d, e} may resemble

$$\frac{1}{2}\,\text{med}\{b, d\} + \frac{1}{2}\,\text{med}\{a, e\}$$

which, for a centered quadratic, would be 5 times as far down (below the peak) as

$$\frac{1}{2}\,\text{med}\{b, d\}\,.$$

Following, rather crudly, the example of the Czech "souslashky na hačky" (consonants with hash marks) like č, š, and ř, we choose to pronounce $\check{4}$ as "foursh", making similar additions of "-sh" to other numerals.

* $\check{5}$ and higher *

In the same spirit, though less violently, if a, b, c, d, e are five successive values, we define five-sh by

$$\check{5} \text{ at } c = \begin{cases} \text{med}(b, c, d) & \text{when}(a-b)(d-e) < 0 \\ \text{med}(a, b, c, d, e) & \text{else} \end{cases}$$

We are now ready to give a recurrent definition, where $n=m+2$ with $m > 3$, by

$$\check{n} = \begin{cases} \check{m} & \text{if the product of the end differences is} < 0 \\ \text{med}\{n \text{ consecutive values of } y\}, & \text{else} \end{cases}$$

Thus for n odd, an apparently peak value will be replaced by the median of exactly 3 adjacent values (for n odd) or of the two adjacent values (for n even).

* $\check{3}$ *

A component somewhat related to the end-value rule and splitting (see later in this section) which is only infrequently different from 3 for noisy inputs is $\check{3}$, defined to produce

$$\text{med}\left\{\text{med}\left\{y_{i-1}, y_i, y_{i+1}\right\}, \text{med}\left\{y_{i-1}, \frac{3y_{i-1}-y_{i-2}}{2}, y_i\right\}, \text{med}\left\{y_{i+1}, \frac{3y_{i+2}-y_{i+1}}{2}, y_i\right\}\right\}$$

as its output. $\check{3}$R does not flatten peaks and valleys quite as much as 3R.

Whether we should also consider

$$\text{med}\left\{y_{i-1}, \frac{3y_{i-1}-y_{i-2}}{2}, y_i, \frac{3y_{i+1}-y_{i+2}}{2}, y_{i+1}\right\}$$

as a -sh-like smoother is unclear.

* $\check{S}$ *

A modification of S, see later in this section, when $\check{3}$R replaces 3R in the fixup phase following splitting, ending, and rejoining.

* $\check{\check{5}}$ *

An untried analog of $\check{3}$ that seems to deserve attention is "$\check{5}$" whose value at t is

$$\text{median}\left\{2y_{t-1}-y_{t-2},\ y_{t-1},\ y_t,\ y_{t+1},\ 2y_{t+1}-y_{t+2}\right\}$$

which is one of the simple smooths that preserves corners formed when all the relevant points lie along two straight lines meeting at a peak (or valley).

* discussion *

The use of -ah smoothing components ($\check{\text{s}}$moothing components, perhaps) thus allows us to have the greater smoothing power of longer medians away from clear peaks or valleys without accepting the degree of erosive action on peaks or valleys that the longer smoothers would ordinarily produce.

We need more comparative experience to know how widely we want to use such components.

Clearly all -ah components (except $\check{3}$ and $\check{5}$) are selectors (when an odd number of values are combined) or semiselectors (when an even number are combined).

A further step in this direction, about whose performance we know even less, fits a straight line to the 4, 5, or more points in question, and applies $\check{4}$, $\check{5}$, etc. to the residuals. (The smooth part of this -ah-ing has then to be combined with the contribution from the straight line.) Whether this step would be for good or bad is hard to say.

* monotonicity *

A simple way to express the fact that a sequence without adjacent ties is (weakly) monotone (globally or over a section) is to require

$$y_t = \text{med}\{y_{t-1}, y_t, y_{t+1}\} \qquad (*)$$

which ensures that $y_{t-1}-y_t$ and $y_{t+1}-y_t$ are not of the same sign, which is equivalent to ensuring that $y_t - y_{t-1}$ and $y_{t+1} - y_t$ are weakly of the same sign.

More generally, a sequence satisfying (*) consists of monotone sections, joined by stretches of two or more equal values. (As we noticed above, this is clearly a consequence of "3R" since (*) says that another "3" will have no effect.)

* C *

If we really want to require (weak) monotonicity, we can ask for (*) for the condensed sequence $\{z_t\}$ in which adjacent ties in $\{y_t\}$ are replaced by a single value. (Thus t in $\{z_t\}$ ordinarily runs through fewer values than t in $\{y_t\}$.) We will later have some use for condensation as a smoothing component, so we plan to identify it by the letter C.

* head banging *

Another way to look at medians of 3 is to suppose that we have formed, somehow, a low sequence $\{L_t\}$ and a high sequence $\{H_t\}$, between whose pairs of values we want the smooth to fall. An easy way to formalize this is to take

$$\text{median}\left\{L_t, y_t, H_t\right\}$$

as the output of a component. This approach generalizes to more-dimensional t (to smoothing in the plane, etc.) (cp. Tukey 1979, Tukey and Tukey 1981), more readily than other simple sequence (one-dimensional-t) interpretations.

* the H component *

If "2" denotes "running means of 2" or "running medians of 2", which are identical, then H = 22 is hanning, definable as

$$\frac{1}{4}y_{t-1} + \frac{1}{2}y_t + \frac{1}{4}y_{t+1}$$

or as

$$\frac{1}{2}\left[\frac{1}{2}y_{t-1} + \frac{1}{2}y_t\right] + \frac{1}{2}\left[\frac{1}{2}y_t + \frac{1}{2}y_{t+1}\right]$$

or as

$$\frac{1}{2}\left(\frac{1}{2}y_{t-1} + \frac{1}{2}y_{t+1}\right) + \frac{1}{2}y_t$$

or in another form to be mentioned in section 9. Except for its linearity, which may be either a pro or a con, its not being even a semiselector, and the failure of H, HH, HHH, ... to stop at any reasonable number of iterations, the formal properties of H are of little help.

In the presence of exotic values, it is a dangerous component to use early in a smoother, particularly because of tenting. Once more robust components have been applied, however, it is often a very useful polishing tool; especially when "local smoothness" is more valued than the "precise values of the smoothed sequence".

* end values and S *

The naive approach to the ends of the input sequence makes use of two forms of a simple idea:

a) shorten the smooth (as in $^{\vee}$ components) when there are only enough values to allow a shorter component (thus at t=2, where only y_1, y_2, y_3 are available symmetrically around t=2, "5" automatically becomes "3") AND, at the very extremes,

b) copying on, where at t+1, we take y_1 as its own smooth.

Stopping with this last is often not good enough. Though we are unclear as to what would be best, we do fairly well with the "end-value-rule" according to which the smooth at t=1 (*mutatis mutandis* at t=n) is

$$E_{(y_1)} = \text{median}\{3z_2 - 2z_3,\ y_1,\ z_2,.\}$$

where z_t is the value of the smooth of $\{y_t\}$ at t.

* splitting *

3R and its relatives tend to leave many pairs of tied adjacent values, particularly 2-mesas and 2-flats, where the tied values are a local maximum or minimum. Some of these are quite all right as they stand, others are clearly exotic. One way of dealing differently with such 2-extremes is *splitting*. Conceptually we divide the sequence between the two values in the tied extreme. Then we apply the end-value rule to the new end of each portion. Now we can reunite the portions, and smooth lightly - - routinely with "3R", exceptionally as desired.

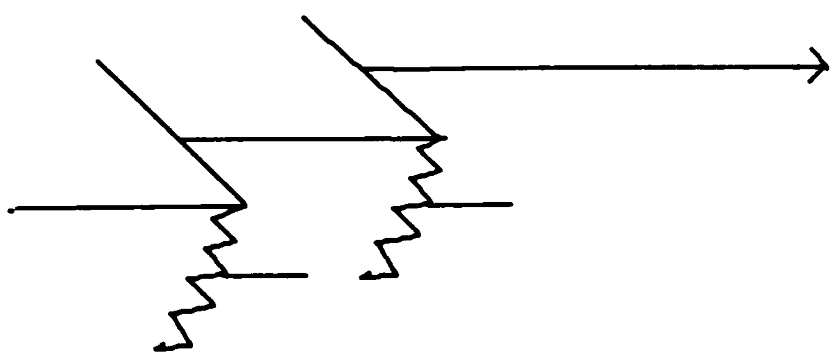
When we want a smooth smooth, "3R" demands something like "S" for "splitting" to follow. Repeating S for the second time is often desirable. (3RSS is a useful work horse.) Indefinite repetition of S can, however, be dangerous, since "zipper-like" action can propagate changes, often unwanted, to indefinite distances.

****** 8. Median-based smoothers -- assembling components ******

To make smoothers out of these components we need to connect them, often in moderately complicated arrangements.

* connectives *

There are only a few simple ways to combine components, particularly resmoothing and reroughing. Resmoothing appears schematically as

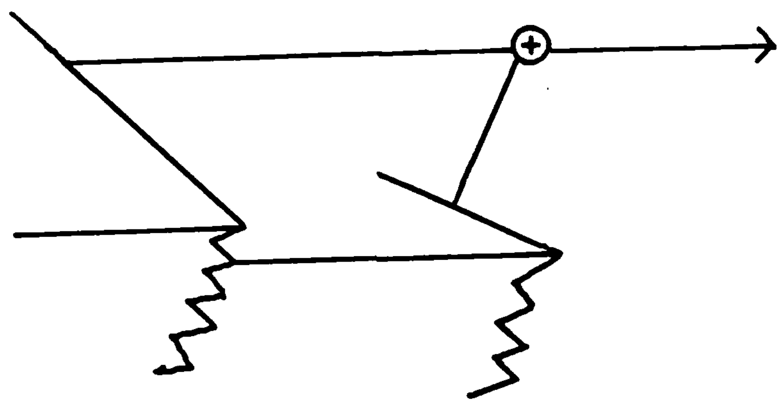


where the divided arrow emits the smooth from its smooth arm and the *rough*, defined by

$$\text{data} \equiv \text{smooth PLUS rough}$$
$$y_i \equiv Sy_i + Ry_i$$

from its rough arm. Resmoothing is most often devoted by simple juxtapostion - - where a separator seems needed we will use a colon.

Reroughing is often denoted by an interposed comma, and appears schematically as



where the smooth of the initial rough is "added back" to the initial smooth. If the two (or more) smoothers in a reroughing configuration are the same, we may, and often do, refer to twicing (thricing, ...).

Indefinite repetition - - repetition "to death" - - is only feasible if the process for any finite sequence comes to a halt after a finite number of steps. Fortunately, as noted above, this does happen for odd-length median smoothings, so that "3R" - - meaning "3 repeated to death" - - is a useful finite process for any finite sequence.

* stranding *

An approach that has been repeatedly suggested as a way to smooth somewhat more vigorously – in a sense down to lower angular frequencies – but seems not to have been tried out extensively is *stranding* (called "slicing" by Gebski and McNeil 1984). Here the original sequence is first divided into k subsequences, each of which contains every kth value from the original sequence. Each of these subsequences is smoothed separately, the results are interleaved to the places from which they came, and further smoothing applied to bring the strands to a common smoothness.

* spacing *

We have subscripted our y's with integers, as if the values came at equally spaced points. What if the spacings are not equal?

For 3, 5, ..., and $\check{5}$, $\check{7}$, ... which only use the ordering of the locations, there seems to be no theoretical reason at all to make any allowance for unequal spacing. Experience seems to confirm this.

For 2, 4, ... and $\check{4}$, $\check{6}$, ..., including H, there would seem to be some theoretical reason to do such things as replacing H by H^*, whose value at t would be

$$\frac{\epsilon}{2(\delta+\epsilon)}\,y_{t-\delta} + \frac{1}{2}y_t + \frac{\delta}{2(\delta+\epsilon)}\,y_{t+\epsilon}$$

which is identical to

$$\frac{1}{2}\left(\frac{\epsilon}{\delta+\epsilon}y_{t-\delta} + \frac{\delta}{\delta+\epsilon}y_{t+\epsilon}\right) + \frac{1}{2}y_t$$

in which the parenthesis can be easily recognized as the linear interpolate from $y_{t-\delta}$ and $y_{t+\epsilon}$ toward t = t. Experience seems so far *not* to have shown such complications to be worthwhile.

For high-performance smoothers (see Section 10) involving - - usually sectionally - - line- or polynomial- fitting it is probably worthwhile to allow for spacing, mainly because of (a) mean-line (i.e. least squares) fitting in the body of the smooth and (b) unsymmetric windows near the ends.

For median-based smoothers, the evidence to date favors "don't bother", as does the simplicity of treating all sequences, however irregularly spaced, as if they were equi-spaced. So we shall say no more about unequal spacing here.

* condensation for global monotonicity *

There are many sequences for which a globally monotone smooth would be UNacceptable. There are others, however, where we might like to reach a monotone result.

The alternating use of 3R - - which enforces monotone sections, joined by flats of length at least 2 - - and C - - which, as we saw, reduces each flat to a single point, thus shortening the length of the sequence - - is a selector. Thus it can be carried on "to death" and the final result will in fact be monotone.

One easy way to keep the notation straight in such a process is to introduce

$$y_{a \text{ to } b} = \text{the common value of } y_a, \cdots, y_t, \cdots, y_b$$

Such *interval subscripts* make going back, say from 3R : C : 3R : C : 3R, which will ordinarily be shorter than the original sequence, to a smoothed sequence defined for each of the original t's quite easy.

* historical account *

It is moderately easy, and moderately accurate, to sort out many resistant smoothers into discrete generations. A reasonable sketch — leaving aside questions of fixups at ends, etc. — follows:

Generation 1. 53H, 35H, and 53QH, both once and twice (Tukey 1971)

Generation 2. 3R, 3RSS, and 3RSSH, both once and twice (Tukey 1977)

Generation 3. High-performance smoothers for long series — based on w-estimates and cosine-arch running linear combinations. (Velleman 1975)

Generation 4. 4323, twice or thrice (Velleman 1975)

Generation 5. 43RSS23RSS (and 43RSS23RSSH) once or twice (Tukey 1974/1985)

Generation 6. $\check{4}\check{3}\check{S}\check{S}2\check{3}R\ \check{S}\check{S}$ or $4\check{3}\check{S}\check{S}2\check{3}R\ \check{S}\check{S}H$, once or twice (Tukey 1974/1985)

Generation 7. $\check{3}R\check{S}\check{S}$ or $\check{3}RS\check{R}$, once or twice (Tukey 1974/1985)

Generation 8. High-performance smoothers using sectionally-fitted lines (See Section 11.)

Generation 9. Forced monotone smoothers, like 3RC3RC 3RC 3RC . . . 3RC = (3RC)R

Generation 10. Swoosh-swoosh smoothers (See Section 9)

Generation 11. Detrivializing smoothers (see Section 10).

Generation 12. Smoothers within bounds (see Section 11).

As of the end of 1975, my recommendations for a reasonable bouquet-or menu — of smoothers from generations 1 to 7 looked like this

Light smoothing (tell twice is true): 3R or $\check{3}R$, once or twice.

Moderate smoothing, preserving breaks: $\check{3}R\check{S}\check{S}$ or 3RSS once or twice.

A little smoother, reduced breaks: 3RSSH or $\check{3}R\check{S}\check{S}H$, once or twice.

Still smoother with breaks gone: first 43RSS23SS, once or twice, then 3 — OR first

$4\check{3}R\check{S}\check{S}2\check{3}R\ \check{S}\check{S}$, once or twice, then 3.

For long series, to reduce harmonic distortion. See Velleman 1975

Note: For clean residuals, always use a twice (or thrice, etc.) smoother, or some other sort of reroughing.

My experience with later generations is not extensive enough to urge me to yet propose an update.

****** 9. Swoosh-swoosh smoothers ******

For some sorts of data, the natural smooth seems to be a sequence of relatively smooth sections connected by points of change. (An extreme form would be a polygonal broken line, where the sections are straight.) To obtain smoothers that give such outputs, we need to supplement the collection of more familiar median-based components, perhaps with those we now illustrate.

* 5-LOCK *

We now introduce one new component, "5-LOCK" by the rule:

(5-LOCK) Any maximal monotone section of length 5 or more, containing at least 3 distinct values, is "locked", so that the next component is not allowed to affect any values in any locked section.

This means that anything long enough to deserve being called a "swoosh" will not be affected by the next component.

Exhibit 1, based on enrollment figures for Yale University (kindly furnished by Professor F. J. Anscombe), shows the effect of applying (read from left to right; treat colons as implying resmoothing)

5-LOCK : 3R : 5-LOCK : 5R : 5-LOCK : 7R : 5-LOCK : –

As a result most of the smooth consists of of monotone sections, either up or down. At most ends, these sections overlap, making a locked peak or locked valley.

In our example there are 7 places where one locked group abuts on another (that moves in the same direction possibly with one unlocked value between), namely:

1815-16, 1823-24, 1830-31, 1846-48, 1866-67, 1884-85, 1895-96

there might also have been gaps, where one or two years belonged to no locked group. We clearly want to consider adding another step – or other steps – to deal with such cases.

* ENDS *

The simplest way to try to deal with the abutting arrows is to introduce "ENDS" in terms of these components:

exhibit 1
Early steps of swoosh-swoosh smoothing
the enrollment in Yale University 1796-1975
(5-LOCKS shown by arrows; unchanged arrows and unchanged values not repeated in later colums; see calculations in exhibit 2 for * = ENDS)

Year	In	3R	5R	7R	*
1796	115				
	123				
	168				
	195				
1800	217				
	217				
	242				
	233				
	200	222			
1805	222	204			
	204				
	196				
	183				
	228				
1810	255				
	305				
	313				
	328				
	350				333
1815	352				333
	298				349
	333				349
	349				
	376				
1820	412				
	407	412			
	481	473			470
	473				470
	459	470			473
1825	470	459,470			473
	454	470			473
	501	474			
	474	496			485
	496				485
1830	502	496			485
	469	485			496
	485				496
	536	514			
	514	536			
1835	572	570			
	570				
	564		570		
	561	564	570		
	608	574	564		

Year	In	3R	5R	7R	*
1840	574		564		
	550				
	537	550			
	559	542,550			
	542	559			550
1845	588	584			550
	584				550
	522		531		558
	517				558
	531				558
1850	555				558
	558				
	605				
	594	605			
	605				
1855	619	605			
	598				
	565	578			
	578				
	641		599		
1860	649	641	617,599		
	599				
	617				
	632				
	644				
1865	682				
	709				699
	699				709
	724				
	736				
1870	755				
	809				
	904				
	955				
	1031				
1875	1051				
	1021	1039			
	1039	1022			
	1022				
	1003	1022			

NOTES: Unchanged columns not repeated. 7R made no changes on this page.

(exhibit 1, continued)

Year	In	3R	5R	7R	*
1880	1037				
	1042				
	1096	1092			1076
	1092				1076
	1086		1092		1076
1885	1076				1092
	1134				
	1245				
	1365				
	1477				
1890	1645				
	1784				
	1969				
	2202				
	2350				
1895	2415				
	2615				
	2645				2624
	2624				2645
	2684				
1900	2542	2684			
	2712				
	2816				
	3142	3138			
	3138	3142			
1905	3806	3605			
	3605				
	3433	3450			
	3450	3433			
	3312				
1910	3282				
	3229	3282			
	3288	3272			
	3272	3288,3272			
	3310	3272			
1915	3267				
	3262				
	2006				
	2554				
	3306				
1920	3820				
	3920				
	4534				
	4461	4534			
	5155				
1925	5316				
	5626	5457			
	5457	5626			
	5788				
	6184				

Year	In	3R	5R	7R	*
1930	5914				
	5615	5631			
	5631	5615			
	5475				
	5362				
1935	5493	5483			
	5483	5493			
	5637				
	5747	5744			
	5744				
1940	5694				
	5454				
	5036	5080			
	5080	5036			
	4056				
1945	3363	4056			
	8733				
	8991				
	9017				
	8519				
1950	7745				
	7688				
	7567				
	7555				
	7369				
1955	7353				
	7664	7488			
	7488				
	7665				
	7793				
1960	8129				
	8221				
	8404				
	8333	8404			
	8614	8539			
1965	8539	8614			
	8654				
	8666	8665			
	8665	8666			
	9385	9214			
1970	9214	9231,9219			
	9231	9219,9231			
	9219	9231			
	9417				
	9661				
1975	9721				

NOTES: Unchanged columns not repeated. 7R made no changes on this page; * made no changes after 1900

C, already discussed, which replaces adjacent tied values by a single value, leaving locks in place (even if they now involve fewer than 5 values).

U, which unlocks one value from each abutting lock

by defining "ENDS" as C then U then 3R, all repeated until there are no more abuttings or gaps. The details for the example are given in exhibit 2, where temporarily removed values are shown by * signs (and are neglected when applying 3R).

* intraswoosh smoothing *

A further step that seems to make good sense is to do some smoothing *within* the monotone stretches — the swooshes. Since no median smoothing component not incorporating averaging has an effect on a monotone stretch, it seems natural to use some form of running means. The simplest choice is of course H, which we write in an unfamiliar form as follows (the "+" and "-" subscripts imply an unwritten 1/2):

$$\Delta y_{t+} = y_{t+1} - y_t$$
$$\Delta^2 y_t = \Delta y_{t+} - \Delta y_{t-} = y_{t+1} - 2y_t + y_{t-1}$$
$$Hy_t = \frac{1}{4} y_{t+1} + \frac{1}{2} y_t + \frac{1}{4} y_{t-1} = y_t + \frac{1}{4} \Delta^2 y_t$$

This form

$$Hy_t = y_t + \frac{1}{4} \Delta^2 y_t$$

makes it easy to always calculate the "correction"

$$+\frac{1}{4} \Delta^2 y_t = \frac{1}{4} y_{t-1} - \frac{1}{2} y_t + \frac{1}{4} y_{t+1} = \frac{1}{4}\left\{\left(y_{t+1} - y_t\right) - \left(y_t - y_{t-1}\right)\right\}$$

and then *apply it or not* as is appropriate.

For our present purposes, we apply it at every t that is not a locked peak or locked valley. Exhibit 3 — shows the calculations for a sample column of 25 years, and the results for the remainder of the sequence.

When we plot the results we get the three panels (which deserve and receive different vertical scales!) of exhibit 4. We see that our smoothing has eliminated the

exhibit 2

The calculations required to apply ENDS to the 8 abuttings in exhibit 1 (5L = relevant part of 5-LOCK)

Panel A (1808-1837)

	Input	*C*	*U*	*3R*	*5L*	*C*	*U*	*3R*	*5L*	*Out*
	183									183
	228									228
1810	258									255
	305									305
	313									313
	328									328
	350					350		333		333
1815	352			350		•		•		333
	298			333		333		349		349
	333					•		•		349
	349									349
	376									376
1820	412	412								412
	412	•								412
	473	473		470						470
	473	•		•						470
	470	470		473						473
1825	470	•		•						473
	470	•		•						473
	474	474								474
	496	496		485						485
	496	•		•						485
1830	496	•		•						485
	485	485		496						496
	485	•		•						496
	514									514
	536									536
1835	664									564
	570	570								570
	570	•								570

exhibit 2 (cont'd)

Panel B
(1840 - 1855)

	Input	*C*	*U*	*3R*	*5L*	*C*	*U*	*3R*	*5L*	*Out*
1840	564									564
	550	550				550				550
	550	•				•				550
	550	•				•				550
	559	559				559	559	550		550
1845	584	584	584	559		•		•		550
	584	•	•	•		•				550
	531	531		531		531	531	555		555
	517	517	517	531		•				555
	531	531		531		•				555
1850	555	555				555				555
	558	558				558				558
	605	605				605				605
	605	•				•				605
	605	•				•				605
1855	605	•				•				605

exhibit 2 (cont'd)

Panel C
(1855 - 1870)

	Input	C	U	3R	5L	Out
1855	605					
	598					598
	578					578
	578					578
	599					599
1860	599					599
	599					599
	617					617
	632					632
	644					644
1865	682					682
	709			699		699
	699			709		709
	724					724
	736					736
1870	755					755

exhibit 2 (cont'd)

Panel D
(1880-1905)

	Input	C	U	3R	5L	Out
1880	1037					1037
	1042					1042
	1092	1092		1076		1076
	1092	1092		1076		1076
	1092	1092		1076		1076
1885	1076			1092		1092
	1134					1134
	1245					1245
	1365					1365
	1477					1477
1890	1645					1645
	1784					1784
	1969					1969
	2202					2202
	2350					2350
1895	2415					2415
	2615					2615
	2645			2624		2624
	2624			2645		2645
	2684	2684				2684
1900	2684	*				2684
	2712					2712
	2816					2816
	3138					3138
	3192					3142
1905	3605					3605

exhibit 3

Intraswoosh smoothing, initial version
(Values in ( ) are locked peaks and locked valleys)

t	y_t	Δy_t	Δy_t^2	$\frac{1}{4}\Delta y_t^2$	Hy_t	Hy_{t+25}	Hy_{t+50}	Hy_{t+75}	Hy_{t+100}	Hy_{t+125}	Hy_{t+150}
1796	115	8	?		(115)*	426	552	819	2574	4053	8988
	123	45	37	9	132	456	538	893	2646	4381	(9017)*
	168	27	-18	-4	164	471	(531)	961	2670	4659	8381
	195	22	-5	-1	194	472	537	1017	2683	5035	7924
1800	217	0	-22	-5	212	473	550	(1051)	2692	5311	7672
	217	25	25	6	223	476	568	1038	2746	5464	7594
	242	-9	-34	-8	(242)	482	593	1026	2871	5625	7512
	233	-11	-2	0	233	485	604	(1022)	3059	5746	7411
	222	-18	-7	-1	221	488	605	1026	3231	(6184)	7353
1805	204	0	18	4	208	493	605	1035	3416	5999	7455
	204	-8	-8	-2	202	496	605	1053	(3467)	5794	7532
	196	-3	-5	-1	195	500	608	1080	3384	5638	7653
	183	45	58	14	(183)	515	614	1092	3332	5483	7845
	228	27	-18	-4	224	537	617	192	3310	(5362)	8068
1810	255	50	23	5	260	557	617	1102	3284	5455	8243
	305	8	-42	-10	295	(564)	617	1157	3280	5527	8359
	313	15	7	1	314	(564)	620	1247	3274	5628	8437
	328	0	-15	-3	325	(564)	632	1363	3272	(5744)	8524
	358	5	5	1	329	(564)	650	1491	3271	(5744)	8606
1815	353	16	11	2	335	563	677	1038	3267	5647	8647
	349	0	-16	-4	345	560	698	1795	2950	5421	8663
	349	0	0	0	349	559	710	1981	(2006)	5162	8802
	349	27	27	6	355	559	724	2182	2605	4802	9079
	376	36	9	2	378	559	737	2331	3247	4152	9220
1820	412	0	-36	-9	403	559	763	2449	3719	(3362)*	9228
	412	58	58	14	426	532	819	2574	4053	7629	9277**

NOTES: y_t is output of Exhibit 1; $\frac{1}{4}\Delta^2 y_t$ is taken to the nearest smaller ($\leqslant$) integer; $Hy_t = y_t + \frac{1}{4}\Delta y_t^2$ except where parenthesized, where $Hy_t = y_t$.

*Only half locked, but treated as locked

**Values of Hy_{t+175} are 9277, 9431, 9615 and (9721)*

exhibit 4

Smoothed Yale enrollment

Panel A

(1796-1866)

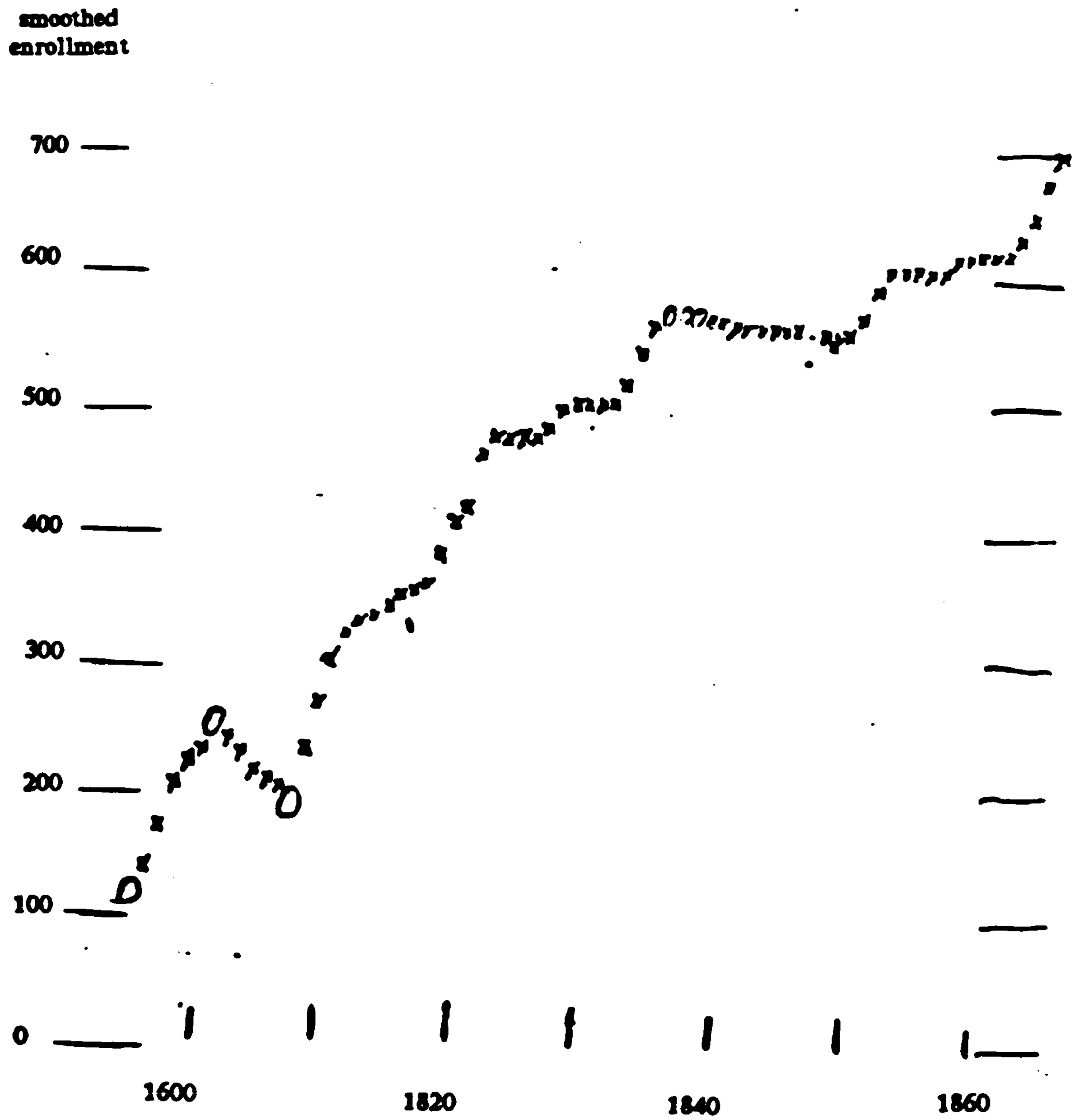


exhibit 4

smoothed enrollment

Smoothed Yale enrollment

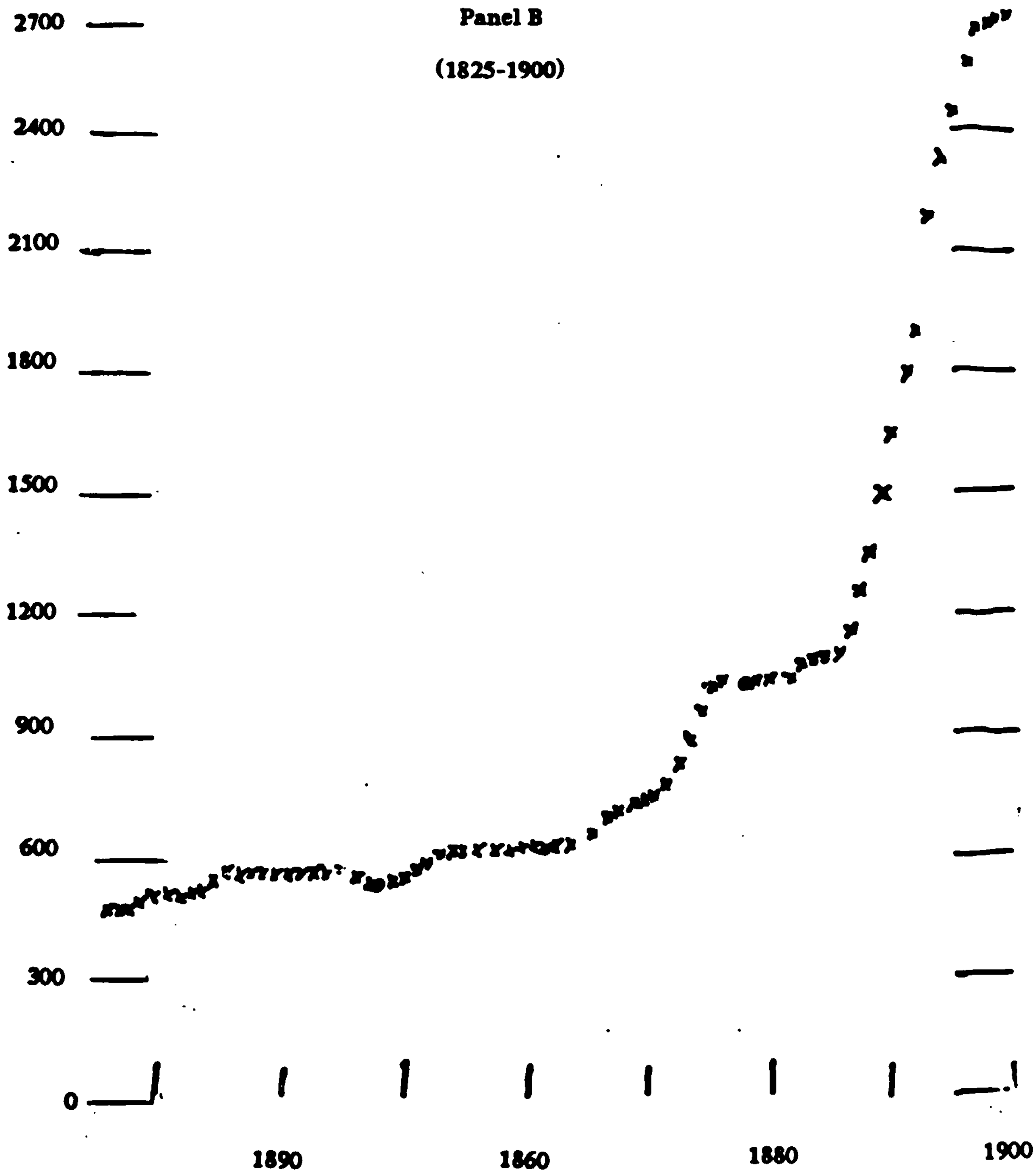
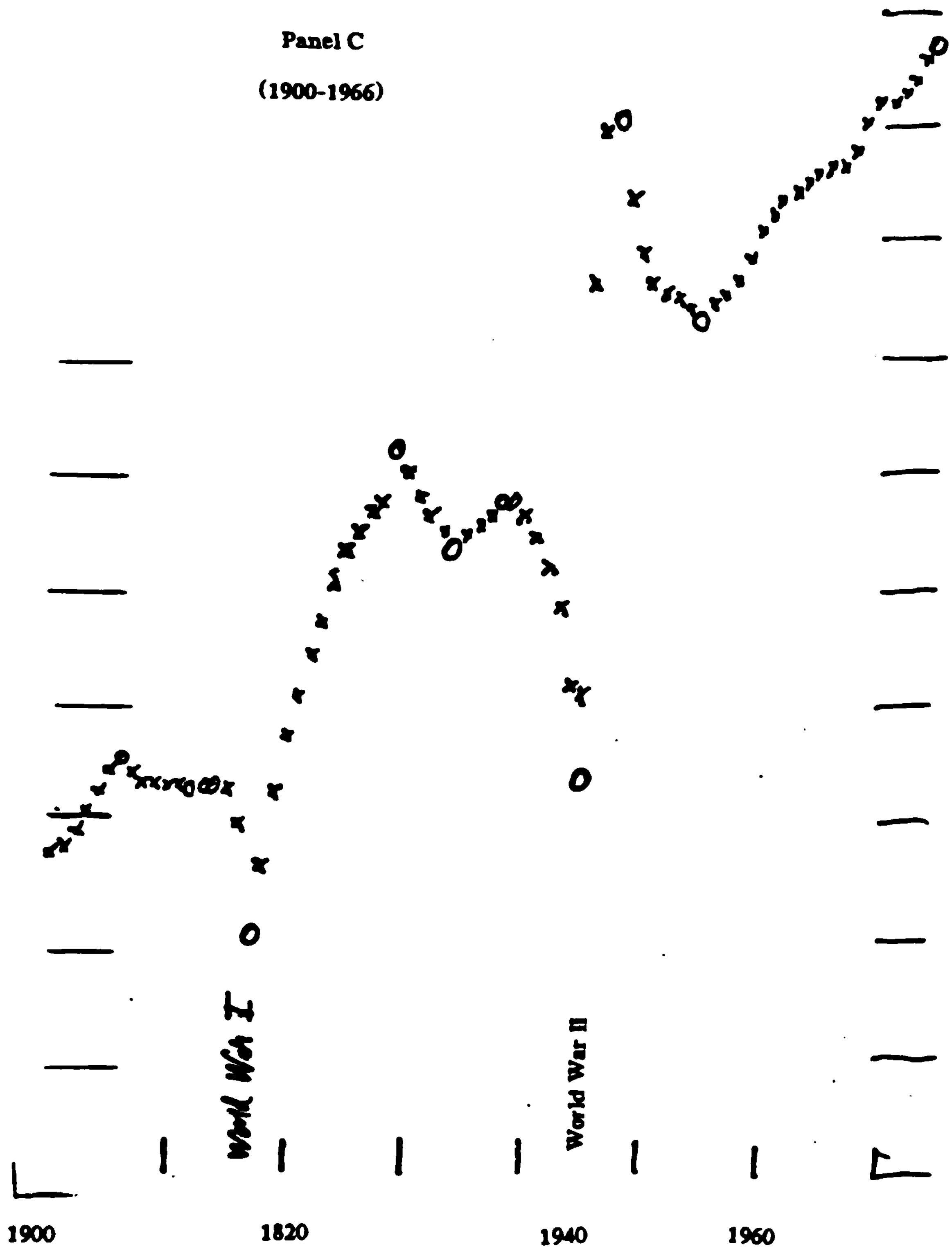


exhibit 4

Smoothed Yale enrollment

Panel C

(1900-1966)



roughnesses that might otherwise distract the eye, without eliminating — or evading — any sudden jumps or relatively narrow peaks or valleys. (The reader may find it interesting to do pure median smooths on the same original data (cp. exhibit 1), plotting the results and comparing them with exhibit 4.

* another revision *

ɔther way to look at the intraswoosh smoothing that we have just done leads to slightly different answers. We can decide to do the H-like smoothing — adding 1/4 of the second difference — at all t's where $\Delta^2 y_t$ is *not unusual*. What evidence might we have for unusualness? Plausibly one of:

a very large value of $\Delta^2 y_t$ compared to what seems natural, OR

a large, but not very large value of $\Delta^2 y_t$ AND a change in direction of monotonicity.

So let us try this in our example. Our first observation — no surprise to any of us — is that $\Delta^2 y_t$'s seem to be larger where the enrollment y_t is larger. Over most of the range of the data sequence the ratio $\Delta^2 y_t / y_t$ seems to behave fairly reasonably. (This may reflect the fact that "first aid" would have urged us to work with logarithms of enrollments.) If we go over to these ratios, and look at (a) only non-zero ratios and (b) only for $t > 1825$ we find a median $|\Delta^2 y_t / y_t|$ of 2.8%.

It is thus plausible to pay special attention to

1) all values of $|\Delta^2 y_t / y_t|$ that are $> 3(2.8\%) = 8.4\%$

2) and those at a turning point that $> 2(2.8\%) = 5.6\%$.

Doing this produces the following special attention list

Years	$\lvert \Delta^2 y_t / y_t \rvert$	Comment
1903 to 1905	10%, 11%, 10%	Fluctuating policy (?)
1916 and 1917	38% and 90%	World War 1
1920 to 1923	11%, 13%, 13%, 14%	Fluctuating policy (?)
1929	11%	Stock market maximum
1934	6%	Minimum
1943	18%	Early World War II
1945 to 1947	180%, 58%, 8%	Return from World War II
1950	9%	Arrest of decline (?)

Before 1825, where the $\lvert \Delta^2 y_t / y_t \rvert$ are generally larger, we must surely single out

1797	30%	???
1808	32%	minimum (why?)

and probably perhaps should include

1802	14%	???
1811	14%	step (why??)
1820-23	9%, 14%, 12%	break (why???)

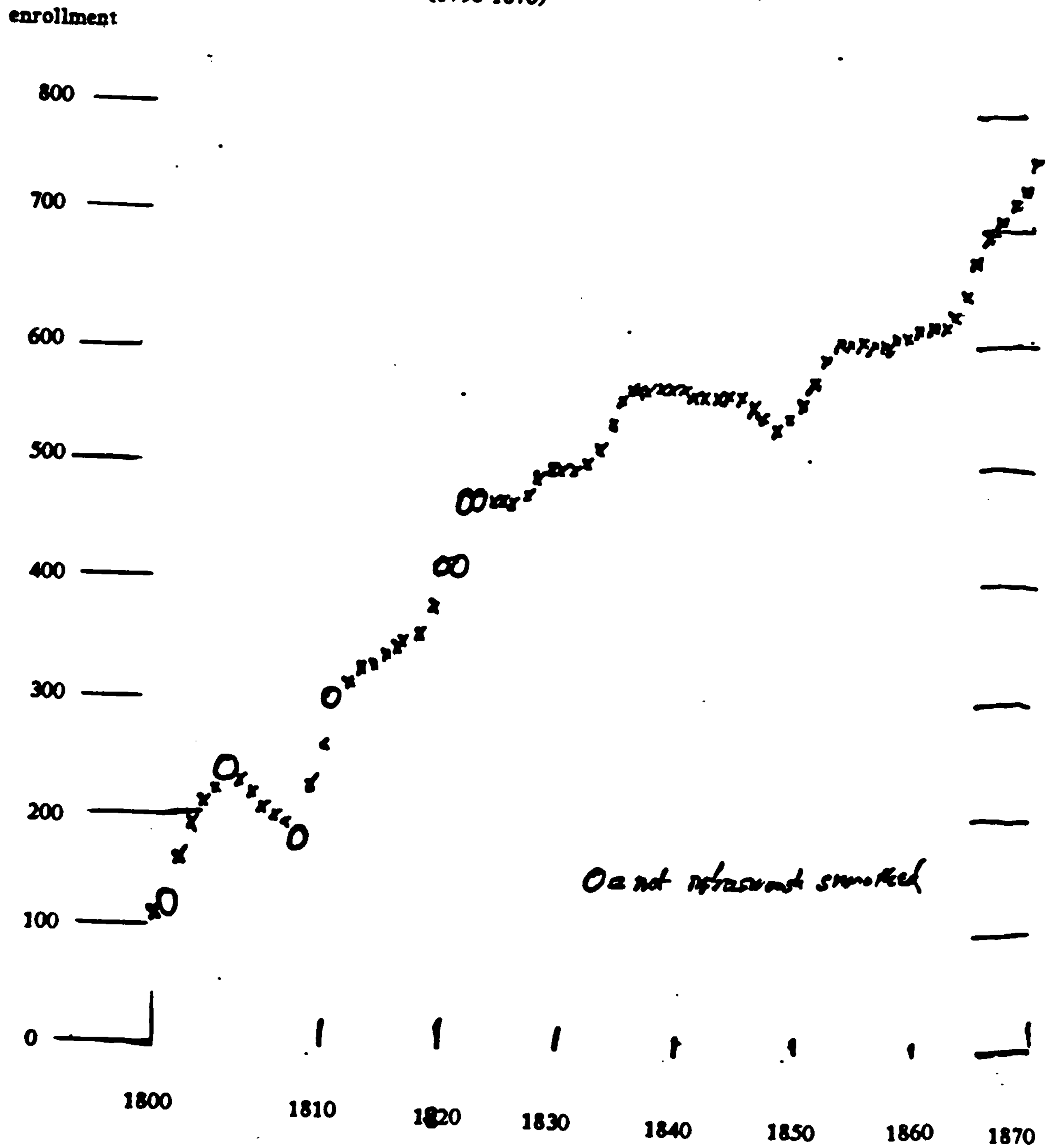
If we leave out all the years thus listed, making the $+\Delta^2 y_t / 4$ adjustment everywhere else, including at the lesser extrema at 1802, 1835-40, 1857-59, 1875, 1877-79, 1938-39, 1948 and 1955, where the size of $\lvert \Delta^2 y_t / y_t \rvert$ does not seem to justify special attention, we get the smooths shown in exhibit 5, which look rather like those of exhibit 4.

However, when we look closely at the points — which have been plotted with a "0" - - where the $\Delta^2 y_t / 4$ adjustment was not applied in exhibit 5 - - we can see that the earlier set (exhibit 4) acts as if some otherwise dull maxima and minima were something special. On the other hand, the later set (exhibit 5) tends to emphasize certain "breaks" as apparently special — e.g. 1821-22, 1905-06, 1916-17, 1922-23, and 1943 and 1945-46. It also indicates disturbance for 1836-38 and 1846-48. Thus the former (exhibit 4) might be more useful if one only wanted a set of smoothed values, without interpretation. And the latter (exhibit 5) would certainly be more

exhibit 5

Revised smoothing of Yale enrollment

Panel A
(1796-1870)



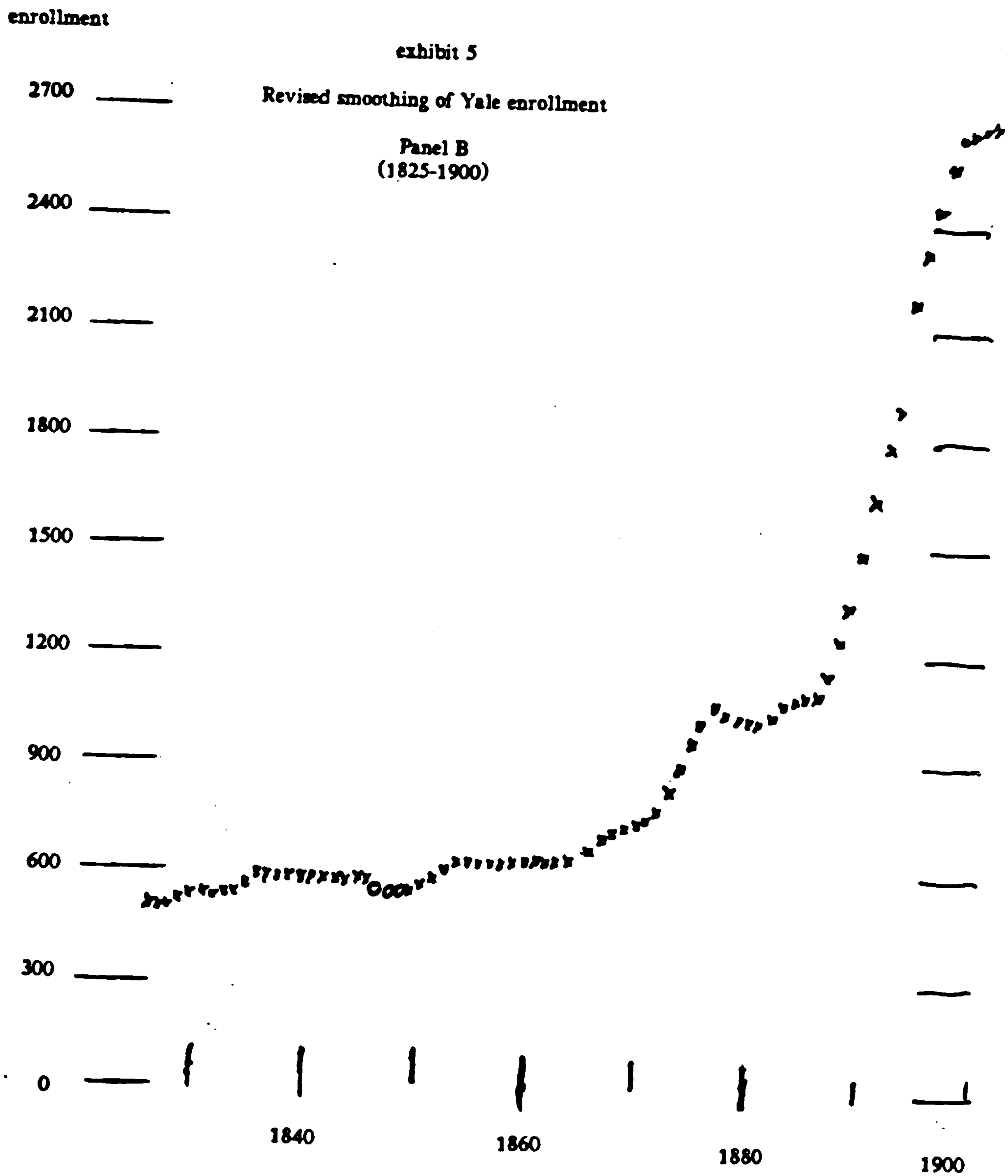
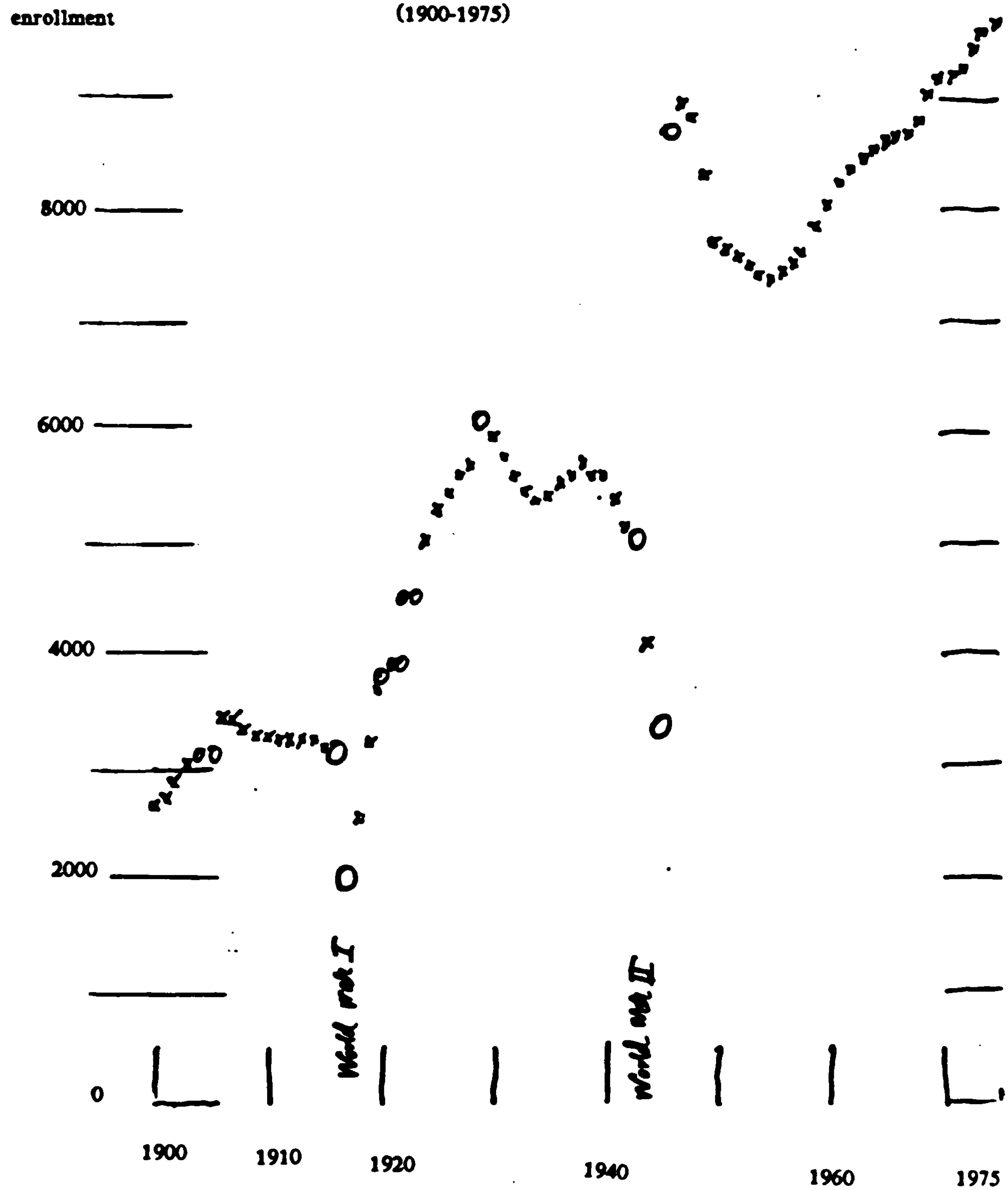

enrollment
exhibit 5
Revised smoothing of Yale enrollment
Panel B
(1825-1900)
2700
2400
2100
1800
1500
1200
900
600
300
0
1840
1860
1880
1900

exhibit 5

Revised smoothing of Yale enrollment

Panel C
(1900-1975)



helpful when one wants a smooth that identifies particular points that appear to respond to either an internal decision or an external event.

This teaches us both (a) that it is not easy to pick out a super-good smoother from a collection of good smoothers — a lot of examples may have to be treated to provide comparisons to establish the kinds and frequencies of relevant differences, since seeing an apparently good performance in one example is awfully little evidence — and (b) that it will often *not* be crucial that we use the absolute best.

* a lesson *

One lesson the potential thinker needs to learn from this example is that differences among relatively good smoothers are often concentrated at relatively infrequent situations.

* choosing the cutoff *

In dealing with "Where should the application of the $.25\Delta^2 y_t$ smooth be cut back to zero?" we have to recognize that most instances of $\Delta^2 y_t = 0$ are the result of three equal values for y_{t-1}, y_t, y_{t+1}. These will probably have come about through the action of 3R, 5R, ... and offer no real evidence of how large $\Delta^2 y_t$ would have been were it not zero.

While in EDA (Tukey 1977) we introduced "starred letter values" where exact zeros count only 1/2 each, it now seems natural to introduce "double-starred letter values" where all exact zeroes (or, conceivably, only exact zeroes of the form 0 - 0, both first differences zero) are excluded from the assessment of typical $|\Delta^2|$. The analysis underlying exhibit 5 was done with

$$\text{cutoff} = 3 \cdot \text{median}^{**}\{|\Delta^2 y_t| / y_t\}$$

and would, for a more simply behaving sequence, have been done with

$$\text{cutoff} = 3 \cdot \text{median}^{**}\{|\Delta^2 y_t|\}$$

Some such choice seems reasonable, at least until we learn more.

Thus G, if we use this notation for the revised version of the limited form of H, is defined by

$$y_t \longrightarrow \begin{cases} y_t, & \text{if } |\Delta^2 y_t| > \text{cutoff} \\ y_t + .25\Delta^2 y_t & \text{else} \end{cases}$$

with "cutoff" as in one of the previous formulas.

Repeated applications of G, as in GG or GGG, have not been excluded, and may prove useful in suitable circumstances.

* suggestions *

Seeing this example obviously generates some interesting possibilities for future study. These seem at the moment to fall into 3 categories:

1) Do we need the step that works on ends of abutting swooshes?

2) What would happen if we used the revised approach on either raw data or much less smoothed data? Need we treat locked peaks and valleys specially?

3) Why not go to LOCK-4 : 3R : LOCK-4 : 5R : — instead of LOCK-5 : 3R : LOCK-5 : 5R : $\cdots$ in the first part of the smoothing?

For the present we leave these questions to the reader.

* should the cutoff be smoothed? *

In a more conventional robustness context, the discontinuity – placed at (3)(M^{**}) in the example above – between applying the $\Delta^2 y_t/4$ correction in its entirety or not at all, would seem to be a lack of smoothness in an amphitheater where lacks of smoothness usually seem to require the payment of a penalty in loss of performance quality. But robust smoothing is not a highly conventional aspect of robustness – in particular, because the various smoothed y_t are not often looked at individually. Moreover it is an area where, if we choose, we can identify, either in a table or in a graphic display, which points are receiving which treatment. We know little about criteria and performance — this leaves us knowing less about the choice between clear discontinuity and more diffuse continuity. Further exploration would be likely to settle this issue, but it is not clear that any great gains are to be made from such a settlement.

* drift in emphasis *

We notice that, while our initial approach to swoosh-swoosh smoothing placed heavy emphasis on the distinction between moving up and moving down, the later versions weaken such emphasis considerably. And the question has been raised — see (2) above — as to whether we could profitably eliminate all reference to "up" and "down". Such changes are not to be thought of as either unlikely or unwise. We are exploring the vast wilderness of the nonlinear — we should expect to follow natural paths, even if they lead us toward an oasis different from the one toward which we started!

****** 10. Detrivialization ******

If we force the evolution of swoosh-swoosh smoothing far enough, we come to a position where we admit, as our basic striving

- to eliminate small rapid wiggles, while preserving both slow changes and large rapid wiggles.

The later modifications of swoosh-swoosh smoothing go a long way in this direction, but it may help other aspects of the reader's thinking to suggest some more general components that may prove useful in this connection.

Let us write $\Delta^2 y_t$ in all our definitions, but let us bear in mind that it may be much better to use $\Delta^2 y_t / y_t$ or $\Delta^2 y_t / y_t^{1/2}$ in appropriate circumstances.

* a class of indicator functions *

The novel characteristic that entered the later subassemblies of swoosh-swoosh smoothing was a "sometimes yes, sometimes no" application of a component according to the value of $|\Delta^2 y_t|$. If we let A stand for the choice of a % and a multiplier, we can define an indicator $I_A(t)$ by

$$I_A(t) = \begin{cases} 1, \text{ when } |\Delta^2 y_t| > (\text{multiplier})(\text{"\% point of } |\Delta^2 y|) \\ 0 \text{ else} \end{cases}$$

with this notation, we can write

$$\begin{aligned} G &= H \quad \text{unless } I_A(t) = 1 \\ &= I \quad \text{else} \end{aligned}$$

where I is the identity, for the application of H except where $|\Delta^2 y_t|$ is large.

We can also, for example, ask about the behavior of

$$3 \text{ unless } I_B(t) = 1$$

separately and in combination with G, where B may equal A, but may involve a different combination of "% point and multiplier.

* rank rather than value *

Another approach would be to calculate all $|\Delta^2|$, sort them, and then act on the smaller ones. Perhaps the 80% smallest? Perhaps the 90% smallest? Perhaps do it 3 times (like 3 hannings) for the 55%, 65% and 75% smallest, respectively. (Much exploration is probably needed.)

Or this could perhaps be combined with the use of indicator functions.

Here then is another "landfall outside the Mediterranian" whose exploration may prove useful -- or uninteresting.

****** 11. "Super smoothers" ******

There are purposes for which a very smooth smooth indeed seems appropriate.

One of these is to prove a somewhat more flexible alternative for both (a) quadratic polynomials and (b) singly-broken lines (monogons) when considering the replacement of an assumed linear dependence by something slightly more general.

Most such smoothers operate by fitting a straight line to a section of the data surrounding the point in question. If there may be exotic points, either this fit has to be robust, or there should be a preliminary application of some other robust smoother.

Almost all smoothers belonging here have one or two tuning constants, to be adjusted to fit each specific situation.

We do not plan to review this class of smoothers with any care, contenting ourselves with identifying some of the most used by name and suggested feelings.

One is W. S. Cleveland's (1979, 1981) *lowess* smoother. This has seen quite a lot of use, and seems to be quite effective. Further detrivialization might help the output's appearance.

Another — or a group of others — comes from Jerome H. Friedman and his coworkers. (See Appendix B, section B1, for further discussion and reference notes.) It is specifically planned for use in re-expression, for example as an important part of the ACE routine.

The procedure for robust spectrum analysis discussed by Martin and Thomson (1982), iterates the two-phase step:

fitting of a simple extrapolator

depending on an estimated spectrum, followed by redescending modification of

innovation = data MINUS extrapolation

While intended to provide a robust spectrum, it does a very good job of eliminating exotic values and should be a near-ideal first step when longer sequences require

robust smoothing.

****** 12. Smoothing within bounds ******

A not infrequent type of problem involves not only values $\{y_t\}$ but measures $\{s_t\}$ for how closely each is likely to be to what it ought to be (unless it is exotic). Doing a good job of responding to this problem will require much more experience than we presently have. Particularly in a piece directed at how to think about such questions, however, there seems to be a place for some tentative explorations.

One very restrictive version would be to look at

$$\mathit{median}\,\{\, y_{t-1},\ y_t - s_t,\ y_t,\ y_t + s_t,\ y_{t+1}\}$$

which always lies in the interval $[y_t - s_t, y_t + s_t]$ and can be thought of as a generalization of head banging.

When we come to iterate such a smooth, we will want to replace y_{t-1}, y_t, y_{t+1} by their respective smooths z_{t-1}, z_t, z_{t+1} but to retain $y_t - s_t$ and $y_t + s_t$. (Similar retentions should occur for the versions that follow!) It can be schematically indicated as

$$\mathit{median}\left\{\begin{array}{c|c} x & (+) \\ xxx & \\ x & (-) \end{array}\right\}$$

where the parenthesized values are multiples of s_t and the columns (<u>not</u> the rows) in the first section correspond to subscripts.

A second, closely related version uses

$$\mathit{median}\,\{y_{t-1}-s_{t-1},\ y_{t-1}+s_{t-1},\ y_t-2s_t,\ y_t-s_t,\ y_t,\ y_t+s_t,\ y_t+2s_t,\ y_{t+1}-s_{t-1},\ y_{t+1}+s_{t+1}\}$$

which can be schematically indicated as

$$\mathit{median}\left\{\begin{array}{c|c} x & (++) \\ xxx & (+) \\ .\,x\,. & \\ xxx & (-) \\ x & (--) \end{array}\right\}$$

This result has to fall in $[y_t-2s_t,\ y_t+2s_t]$, where t is again the schematic horizontal axis, and will often fall in $[y_t-s_t,\ y_t+s_t]$. If we were to iterate, it is not clear which values should come from the current smooth and which from the original data.

Once we have reached this pattern, we can go over to an end-value-like construction, replacing

$$y_{t-1}-s_{t-1} \text{ and } y_{t-1}+s_{t-1}$$

by

$$y_{t-1} \text{ and } 3y_{t-2}-2y_{t-3}$$

and replacing

$$y_{t+1}-s_{t+1} \text{ and } y_{t+1}\ s_{t+1}$$

by

$$y_{t+1} \text{ and } 3y_{t+2}-2y_{t+3}$$

This version seems only likely to be helpful after some initial smoothing, though we must try it out before we understand it.

Firm constraints to $[x_t-s_t,\ x_t+s_t]$ or $[x_t-2s_t,\ x_t+2s_t]$ are likely to be too severe if exotic values, which may be far outside $[y_t-2s_t,\ y_t+2s_t]$ are at all likely; if, for example, we need to face up to measurement fluctuations of estimable size AND to exoticity. In such circumstance we might try such components as

$$\mathit{median}\left\{\begin{array}{ccc:c} & x & & (++) \\ x & . & x & (+) \\ . & x & . & \\ x & . & x & (-) \\ & x & & (--) \end{array}\right\}$$

or

$$\mathit{med}\left\{\begin{array}{ccc:c} x & x & x & (+) \\ . & x & . & \\ x & x & x & (-) \end{array}\right\}$$

which, for each t involved -- each column in the first section -- have more entries

with subscript $\neq t$ than with subscript = t, and, as a result, are not so rigidly restricted.

****** 13. Functionalization ******

We introduced a class of smoothers (at the opening of Section 11) as more flexible alternatives for simple functional forms. Successful fitting of one of the (somewhat?) more flexible forms inevitably leads to the question – motivated by the twin advantages of parsimonious description and of knowing how many constants are effectively being used – "can we do almost as well with a relatively standard parameterized functional form?"

Dealing with this issue requires us to identify some useful functional forms, and consider how to fit them.

Quite a lot of thought and experience tends to leave us with a very few functional forms. Their behavior of most of these is easily describable in terms of their "lodid" or "logarithm of divided difference". This is given, for z a transform of x, and the (z, x) pairs ordered on increasing x, by the combination of the logarithm of the *magnitude* of the divided difference

$$\log \left| \frac{z_{t+1} - z_t}{x_{t+1} - x_t} \right|$$

and the sign of the divided difference.

The proposed standard forms are as follows:

nature	lodid behavior
singly-broken line	two constants, abutting
quadratic (around extremes only)	(first divided-difference linear in x)
exponential	linear in x
power (probably non-integral)	linear in log x

Notice that quadratics are NOT to be considered unless the presence of a maximum or minimum (possibly somewhat outside the data support) is quite certain.

Appropriate techniques for diagnosis and fitting have been described under the name of "smelting" (Tukey, 1981).

****** 14. Approaches to equivariance ******

We often like our data manipulations to have some form — or forms — of compatibility with simple modifications of the input. And then there are times when we are careful to avoid such compatibility.

Most of the techniques of smoothing we have considered here commute with "add a constant" and "multiply by a constant". (The use of $|\Delta^2 y_t| / y_t$ does not commute with "add a constant", however.) They generally do NOT commute, however, with "add a slowly changing function of t", "add a linear function of t" or "multiply by a smoothly changing function of t".

It may help to look at one instance of such non-commuting — so let us take the simplest non-linear component we use often - -"3" - - and three successive values of y_t say 15, 12, and 30.

If we add nothing, we have

"3" applied to 15, 12, 30 is 15 which restores to 15

where "which restores to" means "if we subtract, from the median of the three values (here 15) the value at our *center point* of the added linear function (here identically zero). (After all AC = CA means $C^{-1}AC = A$!)

If we add a linear function of slope 3, say the one with values 100, 103, 106, we may have

"3" applied to 115, 115, 136 is 115 which restores to 12

If we add a linear function of slope -10, say 200, 190, 180, we may have

"3" applied to 215, 202, 210 is 210 which restores to 20

More generally, we get the results in the following table:

slope	restored value
-30	12
-20	12
-18	12
-15	15
-10	20
-22.5	7.5
-5	20
0	15
3	12
5	12
10	12
20	12

where "12" continues unabated for either very large or very small slopes, but a tent-like broken-line dependence takes place between -18 and +3.

Clearly "3"-based smoothers are not equivariant under "addition of a linear function of t".

What can we do about this? Roughly, our choice is either to "forget it" or to both fit and subtract some linear function of t. Clearly the fit can be either global or regional (= segmentwise); clearly we can fit in any of many ways.

The prime versions of "fit and subtract" are the (Cleveland) versions of super smoothers (see Section 11). (It is an interesting question if the Martin and Thomson procedure would be slightly improved by fitting a low-order polynomial either locally or globablly.)

But we can promote equivariance in simpler ways. We might, for example, smooth

$$(y_{t+12} - y_{t-12})/\ 24$$

very severely, and use the resulting value, b, at t = t as a corrective slope for

a) applying "3" to $y_{t-1}+b$, y_t, $y_{t+1}-b$, AND

b) restoring the result.

The point is: We *can* do such things, so we need to *think* about doing them.

These brief indications are included in the hope that they will stimulate both other ideas about, and some comparative study of, smoothing within (or guided by) bounds.

****** 15. A very different application ******

Median smoothers were suggested (pp 631-634) for relating apparent "lines" to background in Tukey 1984j.

16. Conclusions.

Almost all conclusions have to be temporary. We have explored only small patches of the non-linear continent, patches conveniently close to the linear sea and some of its tributary rivers. And we have not been able to help pure exploration appreciably, as yet, by formalizing realistic goals. A few general points, however, seem unlikely to change.

* diversity *

We need to recognize a diversity of aims, and try to meet them with a diversity of smoothers.

* delicacy *

Distinguishing among smoothers that are at least fairly good for the purposes at hand is a delicate matter. Performance for one data set -- or for ten data sets -- may just not be enough to tell us which is to be preferred. Equally, it may not matter that much which one we choose, although it might.

* exoticity *

Techniques which in one way or another treat the exotic differently from the usual are important -- and can play very different roles. (As when resistant smoothers pay minimal attention to exotic values -- but the final phase of swoosh-swoosh smoothing leaves large $|\Delta^2 y_t|$ unadjusted, while smoothing others.)

* experimentation *

Theory is almost certain to consist of numerical experiments, often with stochastically defined inputs. Formula manipulation has so far taught us little.

* erosion *

Some problems will clearly be with us as long as we smooth. One is erosion -- a problem for which we have suggested a variety of palliatives. Reroughing does a lot to minimize the consequences of erosion, but we clearly do not think it does enough -- else why would we have suggested so many modified components where the modification serves to reduce erosion. Moreover, absent erosion, no one might have invented "swoosh-swoosh" smoothing.

Erosion will not go away! But we can expect more and newer devices to eventually reduce its impact still further.

* reader's suggestions *

Suggestions from readers for other useful subjects to be pointed up in this section would be particularly welcome.

Other comments and suggestions are strongly encouraged.

I am happy to thank David Brillinger, David Donoho and Colin Mallows for helpful comments and suggestions, for whose filtering and alteration I take full responsibility.

Appendix A

Antirobust non-linear smoothers and the Beveridge wheat-price series

David Brillinger suggested to me that the famous Beveridge Wheat-price series would be a useful test bed for some newer non-linear smoothers. So some of these were tried out, and, as a consequence, the behavior of the Beveridge series was examined and considered. As detailed below, this series, far from appearing to contain exotic values, seemed to be less irregular at its local extremes than elsewhere. Since such behavior seemed not unreasonable, and might occur in other instances, a smoother was developed which was anti-robust in the sense that the initial steps involved picking out extremes and taking means, with median-smoother components relegated to a minor role, later in the process. The present appendix sets out:

a) the structure of the resulting smoother,

b) the resulting smooth, AND

c) the resulting rough

where all calculations are based on a logarithmic form of the basic series.

****** A1. The character of the Beveridge series ******

A convenient source for the data is pages 623 to 626 of Anderson's book (Anderson 1971). This source gives, as Beveridge did, (i) actual index numbers and (ii) a "trend-free index" obtained by *division* by a 31-year running mean. Since our aim is an additive breakdown, the words "index number" and "division" are trumpet calls toward the taking of logarithms.

It seemed convenient to use logarithms matched at 100 - - so that 100 → 100 and so that the slope at 100 is unity - - this calls for

$$100\ \ln(\text{index}) - 100\ \ln\frac{e}{100} = 100 \cdot \ln\left(\frac{\text{index}}{100}e\right)$$

for which some illustrative values, rounded to integers, as was done with the Beveridge series, are

index	log	index	log
25	-39	100	100
50	31	110	110
60	49	125	122
80	78	150	141
90	89	200	169
100	100	300	210

These illustrative values show rather clearly the qualitative character of the reexpression used.

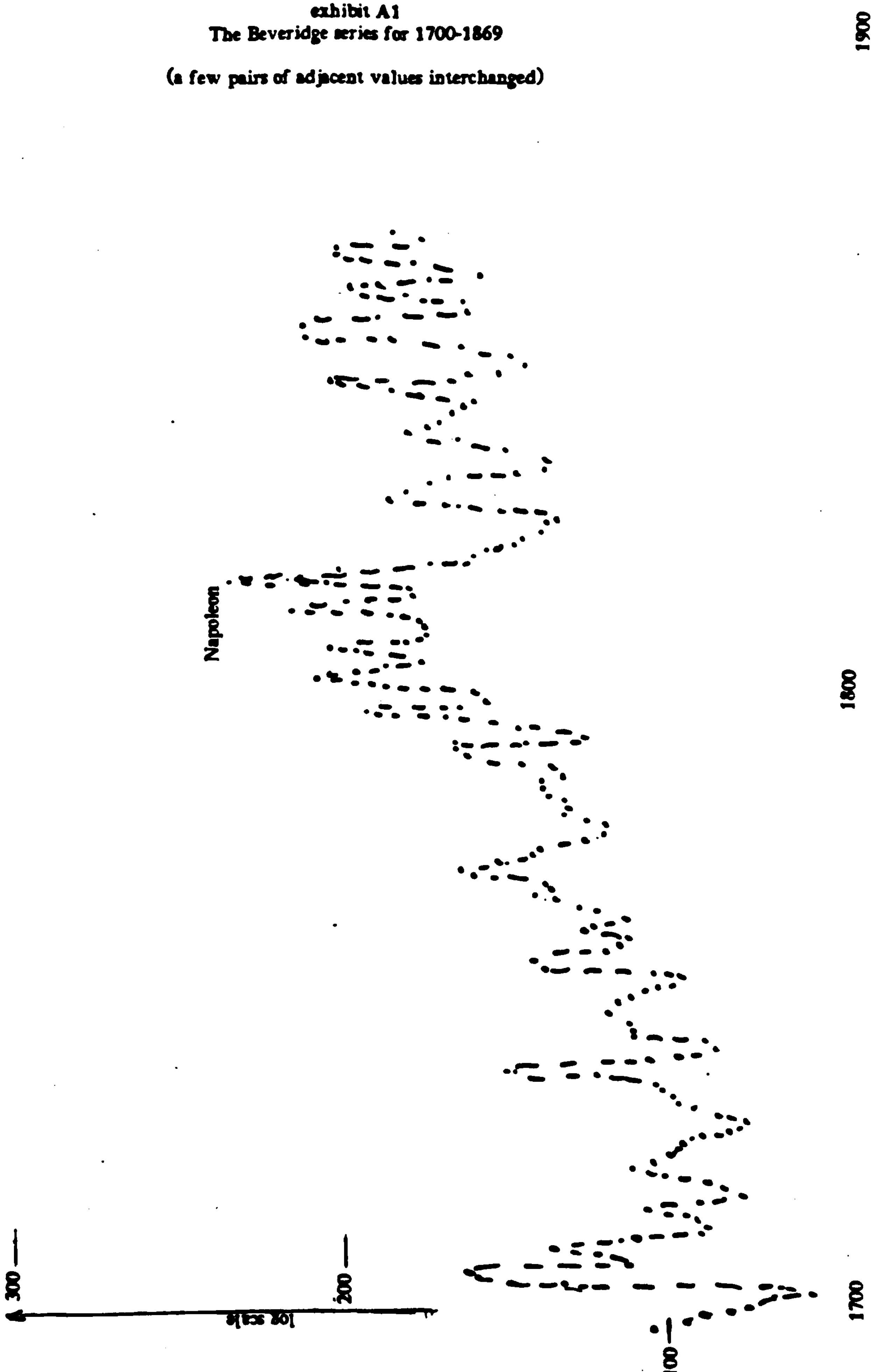
When the original series was modified only by some interchanges of adjacent values, the resulting series for 1700-1869 (the second portion of the series that extends from 1500 to 1869) appears as in exhibit A1. One fairly clear impression that one gains from this plot is a surprising degree of uniformity of size of the ups and downs. (The next most noticeable appearance is the bulge at 1795-1815, contemporaneous with the Napoleonic wars.)

The appearance of this plot is sufficiently well-behaved as to suggest experimentation with smoothers concentrating on local extremes.

exhibit A1

The Beveridge series for 1700-1869

(a few pairs of adjacent values interchanged)



****** 2. The XH3RP smooth ******

The result of a little experimentation, biased toward simplicity and the avoidance of *ad hoc* choices, was a smoother involving the successive application of the following components:

first). No preliminary tinkering, not even adjacent interchanges.

second. X - - identification and selection of all local extremes (centered time for adjacent ties), which must alternate between highs and lows,

third). H - - hanning the selected sequence - - this means linear combinations with weights 1/4, 1/2, 1/4, so that total weight 1/2 goes on one or two lows, and an equal total weight on one or two highs,

fourth). 3R - - meaning medians of 3 applied "to death" (i.e. repeatedly)

fifth). P - - in which short stretches of tied maxima or tied minima - - extrema within the XH3R series - - are replaced by the nearer (in value) of the two adjacent values - - in the XH3R series; here "short stretches" was taken to mean exactly two adjacent values in the selected series tied, the process was iterated as necessary.

It can be argued that the fifth component was slightly *ad hoc*. However, much experience with 3R indicates a real need to do something about tied extremes of length two. Thus our choice does not seem to be seriously *ad hoc*, though it may be too weak.

Exhibit A2 shows the calculations, for all 156 extremes in the original 370-long series.

Exhibits A3 and A4 plot the results for 1500-1700 and 1700-1869, respectively.

exhibit A2

The application of H3RP to the extremes (X) of the Beveridge wheat-price series in 100-matched logarithmic form (only *changed* values shown in 3R and [P] columns)

Year	extreme	X	>	H	3R	[P]
1500	-77	L				
02	-61	H	-90h	-76		
04	-104	L	-79	-91h		
07	-97	H	112h	-100	-96h	[-91h]
09	-121	L	-72	-96h		[91h]
1512	-47	H	-99	-73		
14	-77	L	-54	-65	-73	
16	-61	H	-87	-74		[-73]
18	-97	L	-52	-74h	-74	[-73]
21	-43	H	-93h	-68	-71	
1522	-90	L	-52	-71		
24	-61	H	-93h	77	-71	
25	-97	L	-42h	-70		
31	-24	H	-90	-57	-58	[-67]
34	-83	L	-32h	-58		[-67]
1535h	-51	H	-83	-67		
37	-83	L	-589	-71		
38	-66	H	-80	-13	-71	
40	-77	L	-34	-55h		
45	-2	H	-87	-44h	-49h	
1547	-97	L	-2	-49h	-44h	
51	-2	H	-64	-33		
53	-31	L	14h	-8		[-9]
56	31	H	-37	-3	-8	[-9]
57	-43	L	21	-4	-9	
1562	11	H	-28h	-9		
64	-14	L	17h	2		
65	24	H	-11	6h		
68	-8	L	50	21		
73	76	H	8	42	38	[27]
1575	24	L	52h	38		[27]
77	29	H	22	25h	27	[34h]
78	20	L	34h	27		[34h]
80h	40	H*	29	34h		
82	38	L*	37h	38		

Year	extreme	X	>	H	3R	[P]
1583	42	H	36h	39		
84	35	L	77	56		
86	112	H	41	76	60	
88	47	L	93h	60	65	
90	75	H	55	65		
1591	63	L	75	69		
92	71	H	63	70		
93	63	L	75	69		
1596	109	H	58h	84	70	
1601	54	L	86h	70		
02h	64	H	52h	58	66	[70]
05	51	L	80h	66		[70]
09	97	H	62h	80		
10	74	L*	89	81h	80	
1611	81	H*	76h	79	80	
12	79	L*	80h	80		
13	80	H*	75	77h		
15	71	L	83	77		[77h]
17	86	H	64	75	77	[77h]
1619	57	L	100	78h		
22	104	H	73	93h		
24	89	L	114	101h		
25	114	H	89	101h		
27	89	L	126h	108		
1630	139	H	92h	116	111	[108]
34	96	L	126	111.		[108]
37	113	H	97	105		
39	98	L	108	103		
40	103	H*	99h	101	103	
1641	101	L*	106h	104	102	
42	110	H	92	101	101	
45	83	L	120h	102	101	
49	131	H	65h	98		
50	48	L	130h	89		

NOTES: "h" stands for "and a half". Column "X" is "H" for high extremes, "L" for low ones. When values of "extremes" are very close to one another, "*" is affixed, for later use. Column ">" is "split-mean", containing the arithmetic mean of the "extreme" column values for the previous and following lines; example: -90h = $\frac{1}{2}$((-77) + (-104)). Column "H" is a "line-mean", containing the arithmetic means of the entries in columns "extreme" and "H" in the same line. Columns "3R" show running medians of 3 of the preceding column, repeated as necessary. The "[P]" column shows altered values replacing paired minima or paired maxima (in XH3R values) by the nearer of the adjacent values.

Exhibit A2 (cont'd)

Year	extreme	X	>	H	3R	[P]
1651	130	H	48h	89		
54	49	L	98h	89		
55	67	H*	57h	85h		[89]
56	66	L*	105	85h		[89]
61	143	H	60	101h	91	
1667	54	L	128	91		
74	113	H	69	91		
76	84	L	105	94h	91	
78	97	H	77	87		
82	70	L	96	83		
1684	95	H*	67	81		
85	64	L*	81	72h		[75]
86	67	H*	56h	62	72h	[75]
88	49	L	107h	78		
93	148	H	78h	113		
1695	108	L	149h	129	121	[113]
98	151	H	90h	121		[113]
1702	73	L*	122	97h		
03	93	H*	63h	78	90	[97h]
06	54	L	126h	90		[97h]
09	160	H	81	120h		[115]
11	108	L	146h	127		[115]
1713	133	H	97h	115		
16h	87	L	119h	103		
19	106	H	81h	94		[95]
21	76	L	108h	92	94	[95]
25	111	H	81h	96	95	[94]
1729	87	L*	108h	95		[94]
30	94	H*	80h	87		[93]
32	74	L	98	86	87	[93]
36	102	H	84h	93		
37	95	L	125	110		[106h]
1740	148	H	89h	119	110	[106h]
43	84	L	129	106h		
46	110	H*	96h	103	106h	
47	109	L*	114	111h		[108]
48	118	H*	105	111h		[108]
1750	101	L	115	108		[109h]
51	112	H	97h	105	108	[109h]
54	94	L	125	109h		
1757	138	H	102	120		
60	110	L*	131	120h	120	
61	124	H*	111	117h	120h	
63	112	L	131h	122		
67	139	H	123h	131		

Year	extreme	X	>	H	3R	[P]
1768	135	L	150	142h		
71	161	H	137	149	144	[142h]
73	138	L*	150	144		[142h]
74	139	H*	127h	133		
76	117	L	134h	126		
1777	130	H	116	123	124h	[126]
79	115	L	134	124h		[126]
84	138	H	123h	131		
85	132	L	150	141		
89	162	H	132h	147	143	
1791	133	L	153h	143		
92	145	H*	138h	142	143	
93	144	L*	168	156		
95	191	H	198	169h		
97	152	L	198h	175		
1800	206	H	162h	184		
03	173	L	204	188h	187	
04	202	H	172h	187	188h	
07h	172	L	206h	189		
11	211	H	172h	192		
1813	173	L	222h	198	196	[192]
16	234	H	157h	196		[192]
21	142	L*	189	165h		
22	144	H*	138	141		
23	134	L*	139h	137		[141]
1824	135	H*	132h	134	137	[141]
25	131	L	158h	145		
28	182	H	148h	165h		[163]
29	166	L	179h	173	165	[163]
30	177	H	179	163		
1834	132	L	177	154h	162	[163]
38	177	H	147	162		[163]
40	162	L	172h	167	163h	
41	168	H	159	163h	167	
43	156	L	185	170h		
1846	202	H	198h	175	173h	
49	141	L	206	173h	175	
54	210	H	149h	180		[176]
57	158	L	208h	180		[176]
60	195	H	156	176		[180]
1863	155	L	197	176		[180]
66	199	H	184	181h		
68	173	L	190	181h		
(1869)	(181)	H		(181)		

exhibit A3

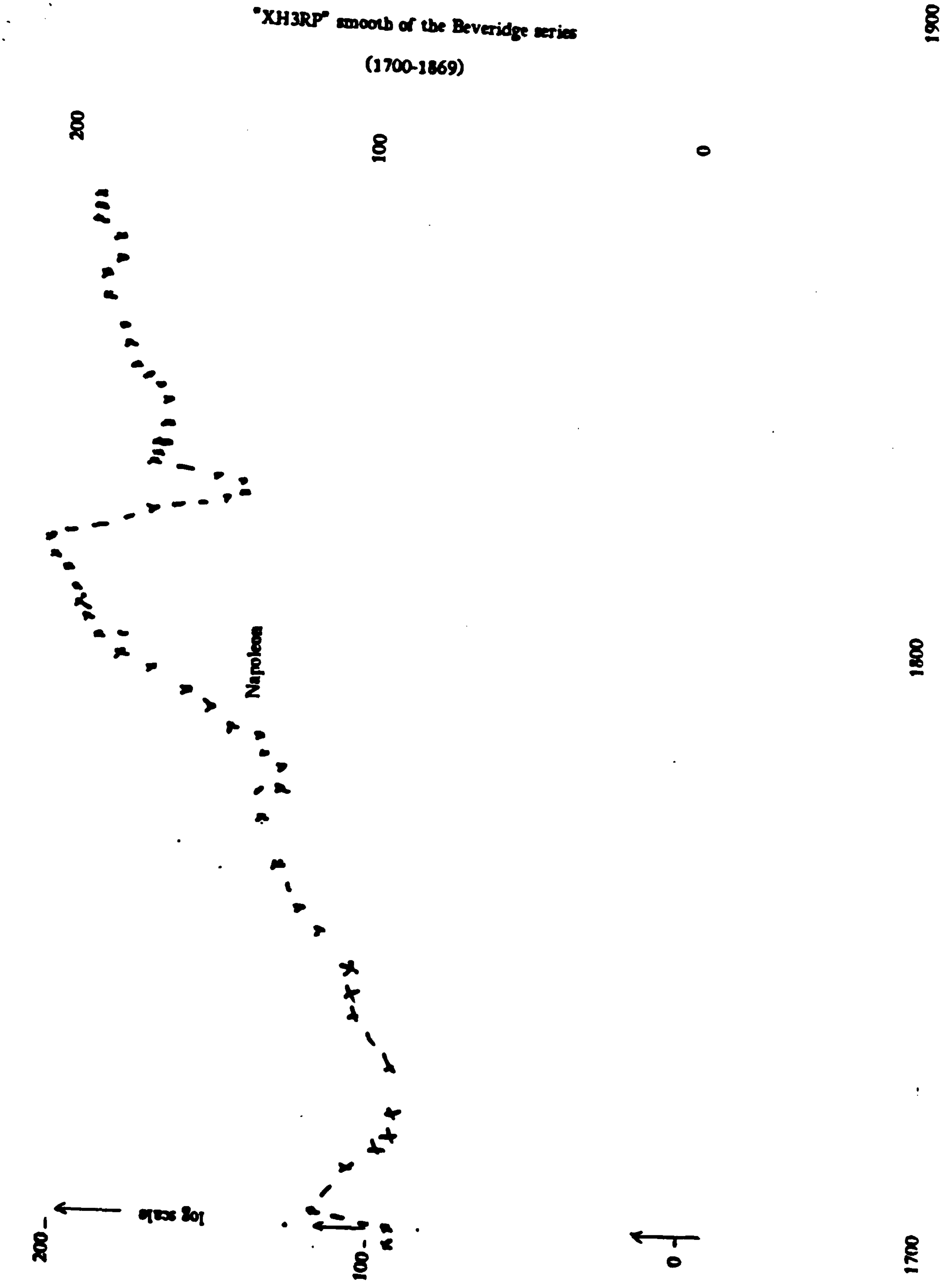
"XH3RP" smooth of the Beveridge series

1500-1700

exhibit A4

"XH3RP" smooth of the Beveridge series

(1700-1869)



****** A3. Possible/plausible modifications ******

If we collect the differences in values between adjacent (unsmoothed) extremes, we get the results in exhibit A5. The distribution seems quite flat in the middle, as it presumably should be (?).

exhibit A5
Stem-and-leaf displays of the peak-to-peak swings in the Beveridge series (log scale)
16 (at 1500 end) and 8 (at 1869 end) omitted
(± 1, ± 2, ± 3 are underscored)

Stem	1500's	1000's	1700's	1800's	Total (cumulative)
10*			6		1
9*	5	9			2 (3)
8*	4	2			2 (5)
7*	3457				4 (9)
6*	2			19	3 (12)
5*	49	07	34	1	7 (19)
4*	6	368	47	456	9 (25)
3*	28		05	79	6 (34)
2*	289	55	035678	9	13 (47)
1*	267	057	123499	15	14 (61)
0*	457	13579	17	126	13 (74)
-0*	979	922	76411	4	13 (75)
-1*	221	8753	8752	6520	12 (62)
-2*	9511	9553	984320	6	15 (50)
-3*	66220	1	990	830	12 (35)
-4*	73	30	6	60	7 (23)
-5*	952		2	2	5 (14)
-6*	5	0	4	1	4 (9)
-7*	4	8			2 (5)
-8*		3			1 (3)
-9*	5			2	2

If we decide to try expunging extrema which contribute to a difference of only ± 1 or ± 2 we get changes in 8 portions of the series (3 rather near each other in 1605-1614, 2 in 1773-1793) as calculated in exhibit A6 and displayed in exhibits A7 and A8. It is interesting to note that, in every case, the expunged extremes involved adjacent years.

exhibit A6

Elimination of extrema contributing to peak-to-peak shifts of ± 1 or ± 2
(blanks in ">" column indicate extremes expunges,
[P] gives final smoothed extremes, (A4) gives last two columns of exhibit A4)

Year	extreme	>	H	3R	[P]	(A4)
73	76	8	42	38	[28]	(38) [25h]
1575	24	52h	38	28		(38) [25h]
77	29	22	25h	28		(27h) [34h]
78	20	35h	28	35		(27h) [34h]
80h	40					(34h)
82	38					(38)
83	42	27h	35			(39)
84	35		56			(56)
• • • •						
1605	51	80	66		[70]	(66) [70]
09	97	62h	80	79	[75h]	(80)
10	74	84	79		[75h]	(80)
11	81					(80)
12	79				[75h]	(80)
13	80					(79h)
15	71	80	75h		[78h]	(77h) [77h]
17	86	64	75	75h	[78h]	(77h) [77h]
19	57		78h			(78h)
• • • •						
1630	139	92h	116	111	[108]	(111) [108]
34	96	126	111		[108]	(111) [108]
37	113	97	105			(105)
39	98	111h	105			(103)
40	103					(103)
41	101					(102)
42	110	90h	100	102		(102)
45	83	120h	102	100		(102)
49	131	65h	98			(98)
1650	48	130h	89			(89)
51	130	48h	89			(89)
54	49	136h	93		[91]	(89)
55	67					(85h) [89]
56	66					(85h) [89]
61	143	51h	98	93	[91]	(91)
67	54	128	91			(91)
74	113	69	91			(91)
• • • •						
1736	102	89h	93			(93)
37	95	125	110		[108h]	(110) [106h]
40	148	89h	119	110	[108h]	(110) [106h]
43	84	133	108h			(106h) [110]
46	110					(106h) [110]
47	109					(111h) [108]
48	118	92h	105	108		(111h)[108]
50	101	115	108	105	108	(108) [109h]
51	112	97h	105	108		(108) [109h]
54	94	125	109h			(109h)

Year	extreme	>	H	3R	[P]	(A4)
1767	139	123h	131			(131)
68	135	150	142h			[131](142h)
71	161	126	143	142h	[131]	(144) [142h]
73	138					(144) [142h]
74	139					(133)
76	117	145h	131			(126)
77	130	116	123	124h	[131]	(124h) [126]
79	115	134	124h		[131]	(124h) [126]
84	138	123h	131			(131)
85	132	150	141			(141)
89	162	132h	147			(143)
91	133	176h	155			(143)
92	145					(143)
93	144					(156)
95	191	142h	167			(169h)
97	152	198h	175			(175)
1800	206	162h	184			(184)
• • • •						
1807h	172	206h	189			(189)
11	211	172h	192			(192)
13	173	22h	198	193	[192]	(196) [192]
16	134	152	193		[192]	(196) [192]
21	142					(165h)
22	144					(141)
23	134					(137) [141]
24	135					(137) [141]
25	131	208	169h			(145)
28	182	148h	165	169h		(165)
29	166	179h	173	165		(165)
1830	177	149	163			(163)

exhibit A7

Effects of modification
(1560-1670)

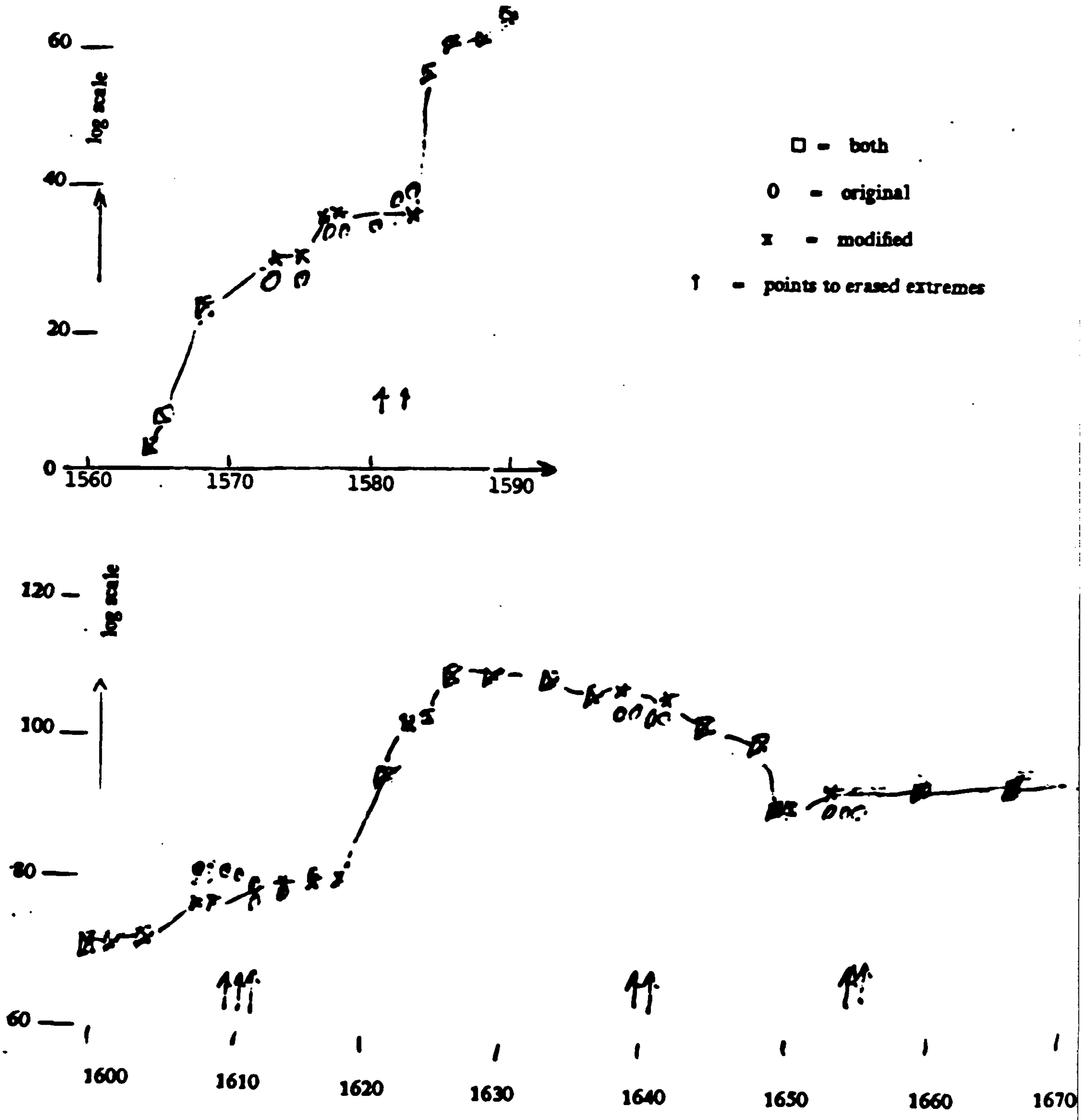
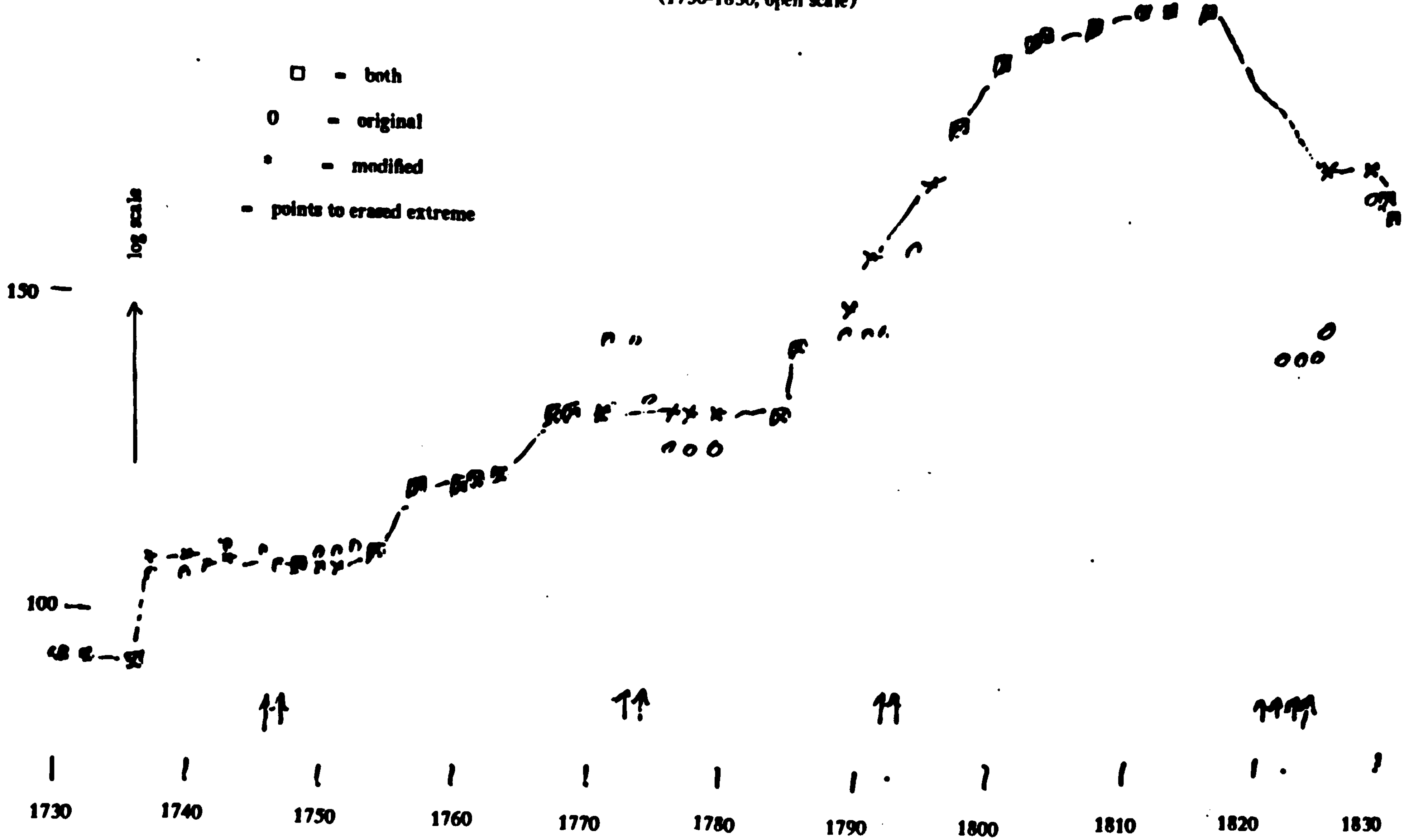


exhibit A8

The effects of modification
(1730-1830, open scale)



****** A4. Smoothing the peak-to-peak changes ******

We have looked at the general trend, but not yet at the degree of oscillation. Exhibit A9 smooths the absolute values of the peak-to-peak swings, the result is plotted in exhibit A10.

We can inquire into the reasonability of our omission of the ± 1 and ± 2 swings by noting their effect on the smoothed peak-to-peak values. Calculations are exhibit A11, where the one ± 3 is also excised and the results in exhibit A12. All the deep valleys in exhibit A10 have disappeared; most of the changes have had such an effect. On the whole the elimination of the ± 1, ± 2, and ± 3 changes seems to have been helpful.

There is some reason for suspecting that "peak-to-peak" assessment of swing is less stable than other assessments might be. To this end, exhibit A13 shows a smooth of

|peak of one kind MINUS median of adjacent peaks of the other|

which is otherwise comparable to the first section of exhibit A12. It seems that this assessment may be more stable, but not by enough to urge us to follow through for the other sections at this point. (Ratios of max to min are: 62/9 = 6.9 in A12 and 64/15h = 4.2 in A13.)

****** A5. Detrivialization to smoothness ******

Turning back to the modified XH3RP smooth (cp. exhibits A3, A4, A7, A8) which is intended to portray "typical" behavior, we easily see that the greatest improvement in overall quality is likely to come from the removal of distracting wiggles. To this end we can apply detrivializers.

exhibit A9

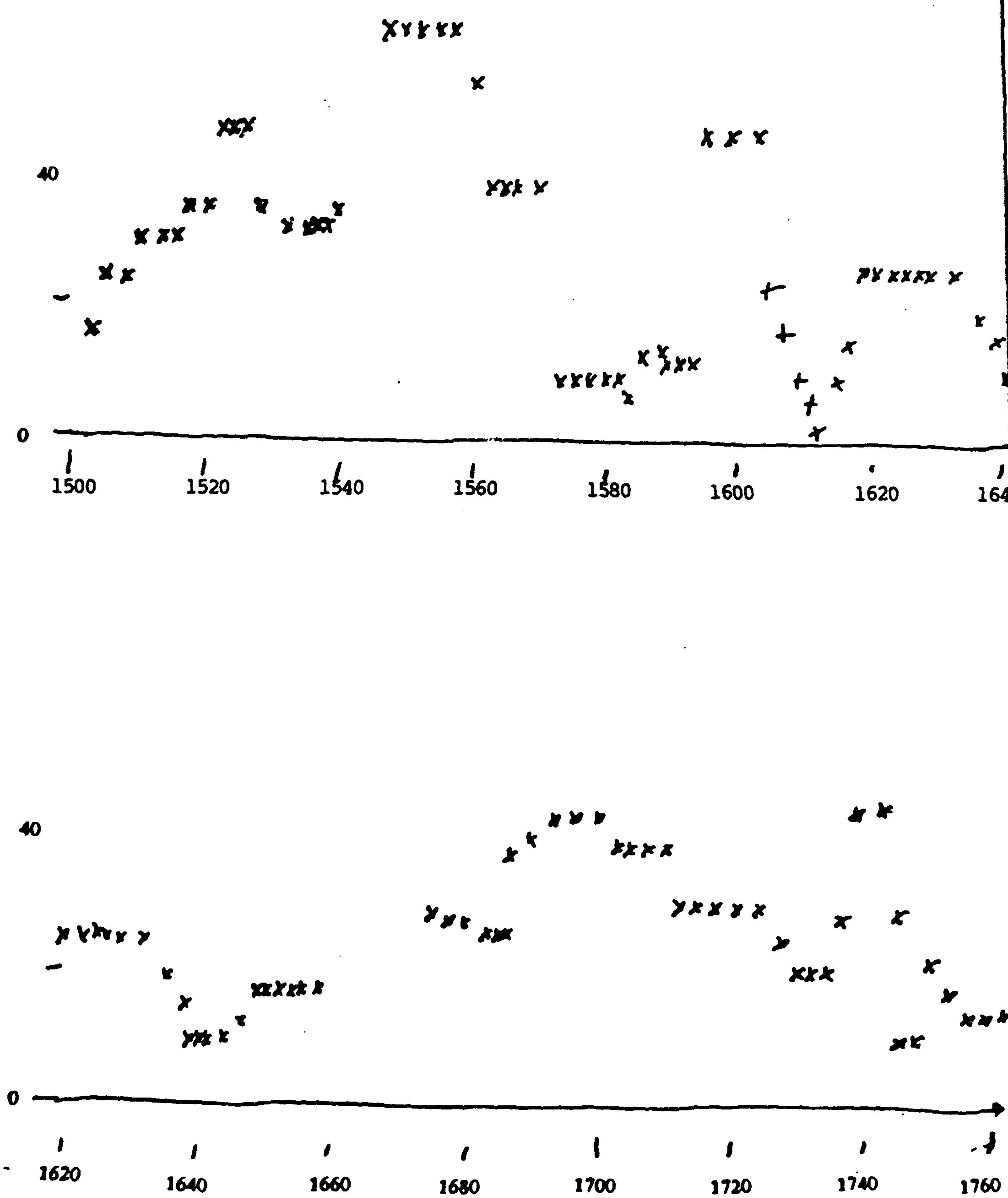
The swings from peak to peak (not counting either 1500 or 1869 as a peak)

swings	(*)	and their smoothing date	swings	(*)	and their smoothing date	swings	(*)	and their smoothing date	swings	(*)	and their smoothing date
-43	16	1503	-7	23	1604	-52	39	1710	-92	22	1818
+7	24	05	+46	17	07	+25	30	1712	+2*	10	21h
-24	24	08	-23	10	09h	-46	30	14	-10	4	22h
+74	30	10	+7	7	10h	+19	30	18	+1*	4	22h
-30	30	13	-2*	2	1611h	-30	30	20	-4	4	1824h
+16	30	1515	+1*	2	12h	+35	30	27	+57	16	1826
-36	36	17	-9	9	14	-24	24	1729	-16	16	28h
+54	47	20	+15	15	16	+7	20	29h	+11	16	29h
-47	47	21h	-29	25	18	-20	20	31	-45	16	32
+29	47	23	+57	25	1620	+28	20	34	+45	16	1839
-36	47	1524h	-25	25	1623	-7	28	36h	-15	16	1839
+73	36	28	+25	25	24h	+53	44	1738	+6	16	42
-59	32	32	-25	25	26	-64	44	42	-12	16	42
+32	32	32	+50	25	28h	+26	26	44	+46	46	44
+32	32	35	-43	25	32	-1*	9	46h	-61	61	48
-32	32	36	+19	19	1636	+9	9	47h	+69	61	1852
+17	32	1537h	-15	15	38	-17	11	50h	-52	-52	56
-11	35	39	+5	9	39h	-18	17	52	+37	-40	58
+75	41	42	-2	9	40h	44	14	51	-40	-40	62
-95	62	46	+9	9	41h	-28	14	58	+44	40	64
+95	62	49	-27	9	1644	+14	14	1760h	-26	-26	1867
-29	62	1552	+48	13	47	-12	14	1762			
+62	62	54	-13	18	49h	+27	14	66			
-74	62	56h	+12	18	50h	-4	23	67h			
54	54	60	-81	18	52	+26	23	70			
-25	38	63	+18	18	1654h	-23	23	1772			
+38	38	1569h	-1	18	55h	+1*	22	73h			
-32	38	66	+77	18	58	-22	21	75			
+84	38	70	-89	59	64	+13	15	76h			
-52	59	74	+59	59	70	-15	15	78			
+5	9	76	-29	29	1675	+23	15	1782			
-9	9	1577h	+13	27	77	-6	23	84h			
+20	9	79	-27	27	80	+30	23	87			
-2	9	81	+25	25	83	-29	29	90			
+4	9	82h	-31	25	84h	12	29	91h			
-7	9	83h	+3*	25	1685h	-1*	39	94			
+77	13	1585	-18	37	87	+47	39	94			
-65	13	87	+99	40	90	-39	39	96			
+28	12	89	-40	43	94	+54	39	98			
-12	12	90h	+43	43	96h	-33	39	1802			
+12	12	91h	-78	43	1700	+29	36	03h			
-12	12	1592h	+26	39	02h	-30	36	06			
+46	46	94	-39	39	04	+39	38	09			
-55	46	98	+106	39	08	-38	38	12			
+10	46	1602	-52	39	1710	+61	38	1814			

*3RSS, twice applied to absolute value of swings.

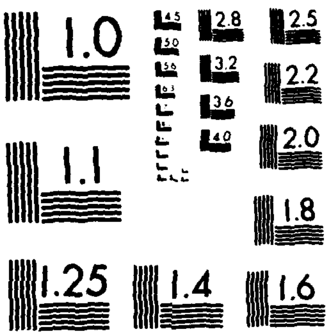
exhibit A10

The smooth of peak-to-peak
(from exhibit A9; 1st 2 panels)



AD-A172 738 THINKING ABOUT NON-LINEAR SMOOTHERS(U) PRINCETON UNIV NJ DEPT OF STATISTICS J W TUKEY MAY 86 TR-291-SER-2 ARO-23360.4-MA DAAL03-86-K-0073 2/2

UNCLASSIFIED F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

exhibit A10 (cont'd)

The smooth of A9 for 1760-1869

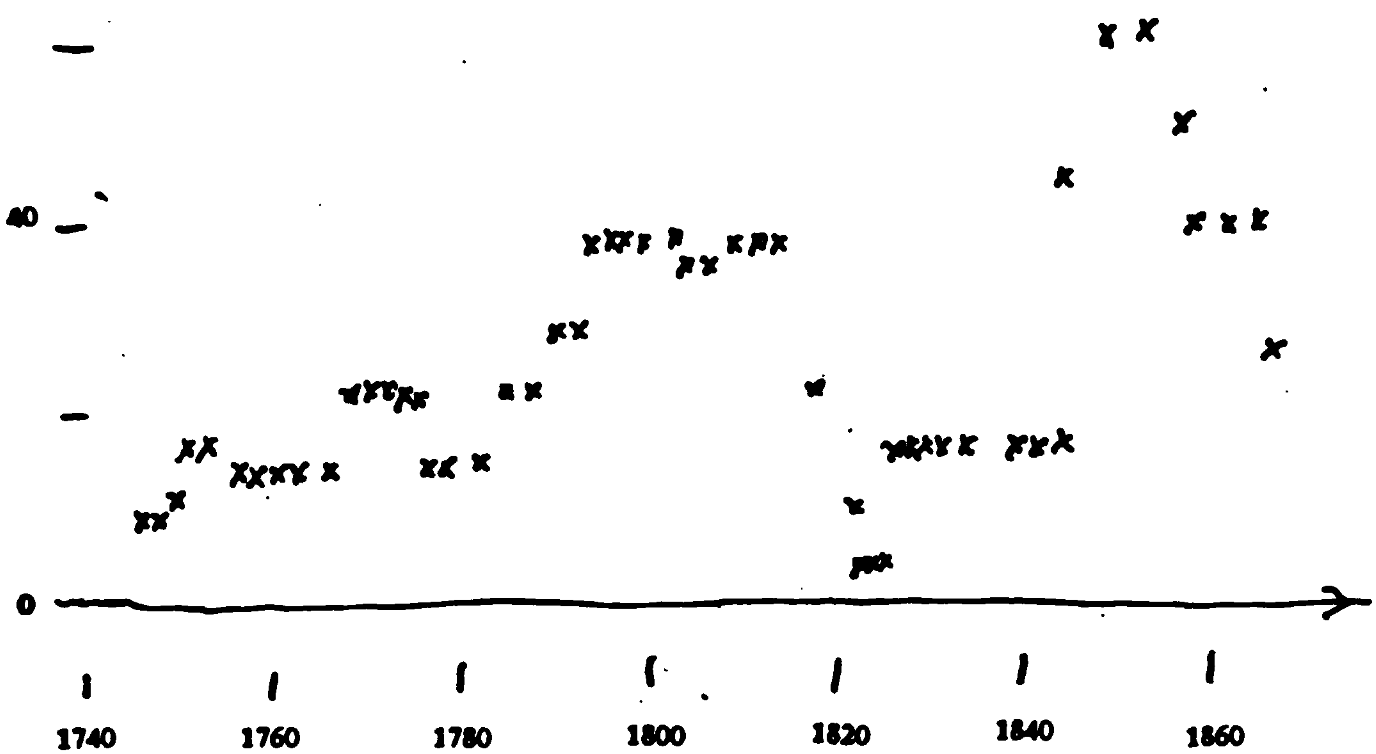


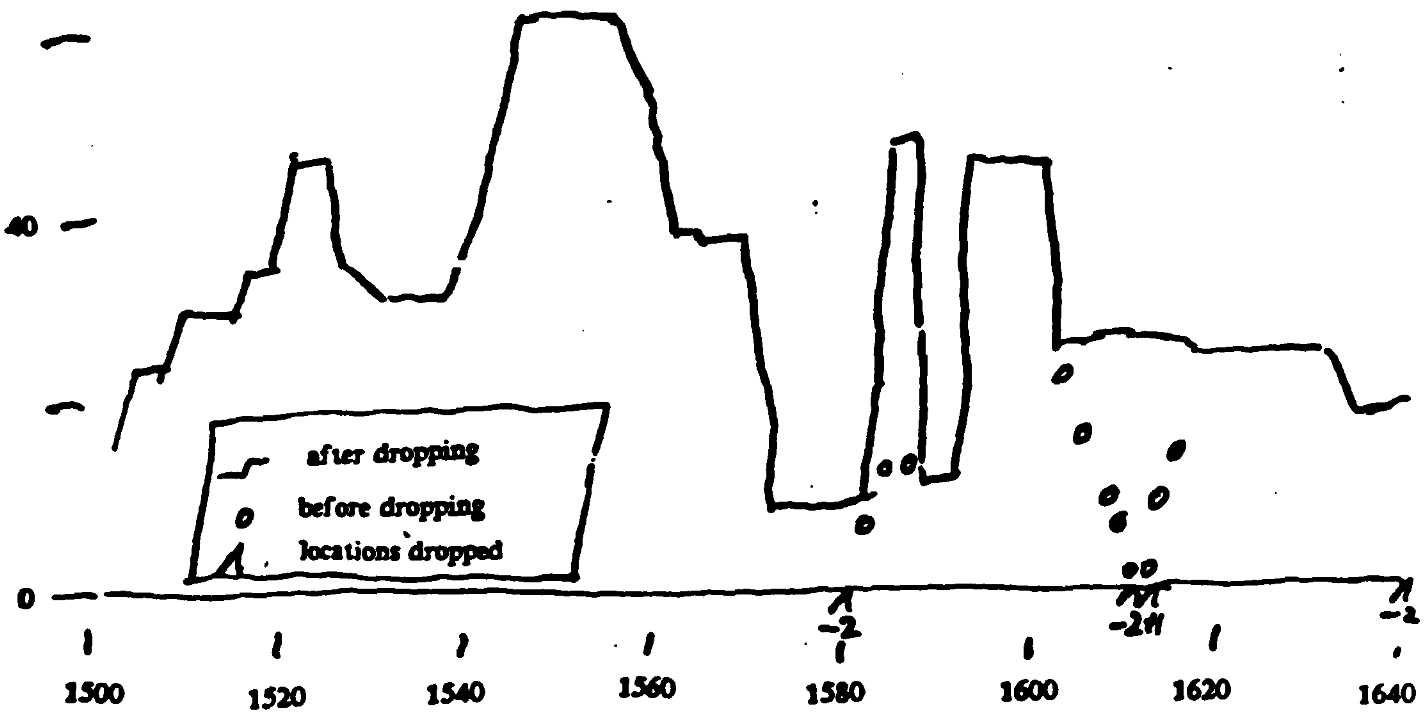
exhibit A11

Modified swings and smooths

swing	smooth	year	swing	smooth	year
84	38	1570	+27	14	1766
-52	9	74	-4	26	68h
-9	9	77h	+26	26	70
→+23	9	80	-44	26	74
-7	32	85h	+13	15	76h
77	48	85	-15	15	78
-65	48	87	+23	15	82
28	28	89	-6	23	84h
12	12	1590h	+30	29	87
		....	-29	30	30
+10	46	1602	→+58	39	93
-7	26	04	39	39	96
+46	26	07	54	39	98
→-26	26	12	29	36	1803h
+15	26	16	30	36	06
-29	25	18	39	38	09
+57	25	1620	-38	39	12
+19	19	1636	+61	61	14
-15	19	39	→-103	61	20
→-27	19	44	+51	61	26
+48	19	47	16	61	28h
-13	18	49h	+11	45	29h
12	18	50h	-45	23	34
-81	18	1652	-45	23	34
		....	-15	23	39
-29	29	1675	6	16	40h
+13	27	77	-12	16	1842
-27	27	80			
+25	27	83			
-46	27	86			
99	43	90			
-40	43	94			
+43	43	96			
+9	43	1700			
		....			
+28	20	1734			
-7	28	36h			
+53	44	38			
-64	44	42			
→34	34	46			
-17	17	49			
11	17	50h			
-18	17	52			
44	17	56			
-28	14	1758			
		....			

exhibit A12

Smooth after dropping
"extreme" next to small changes



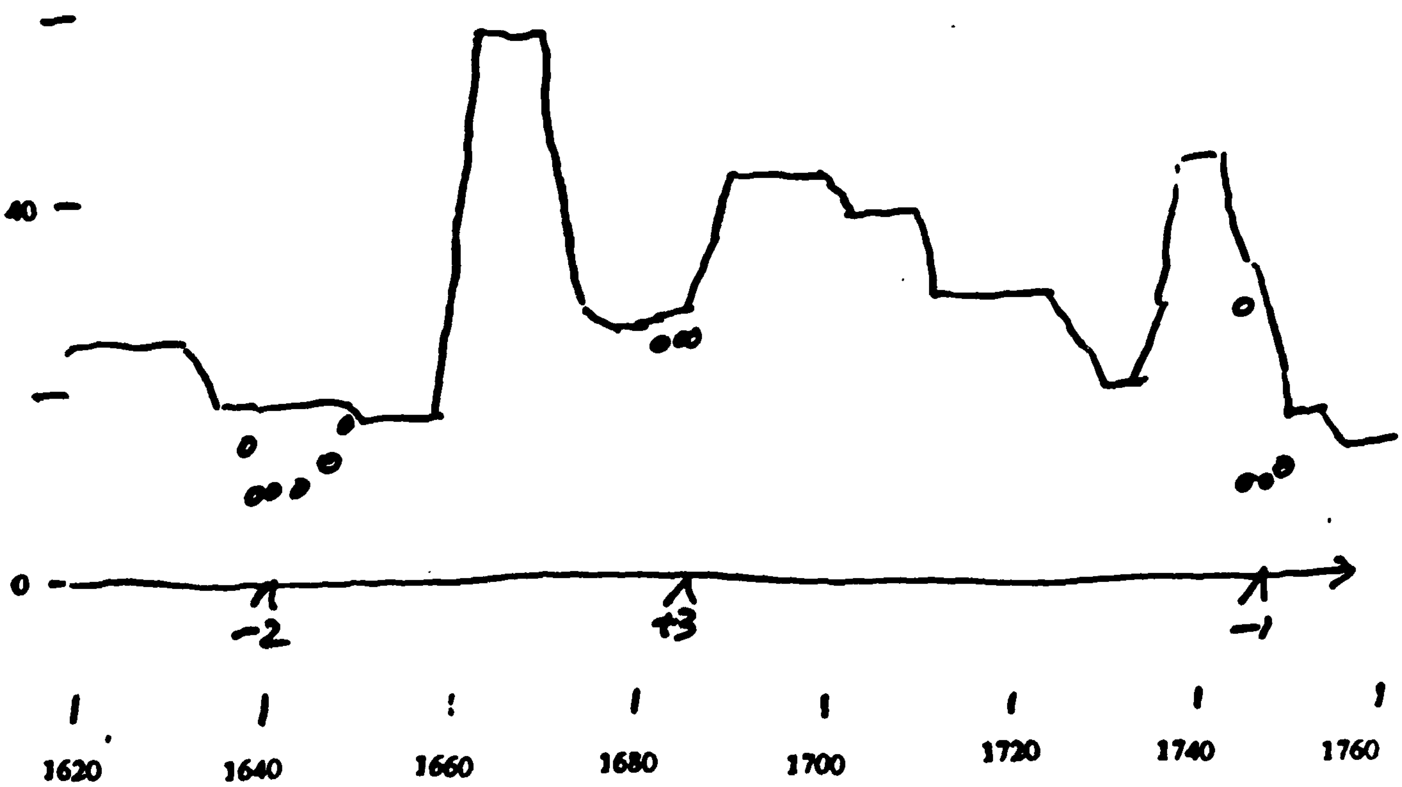


exhibit A12 (cont'd)

Smooth after dropping
(third panel, 1760-1864)

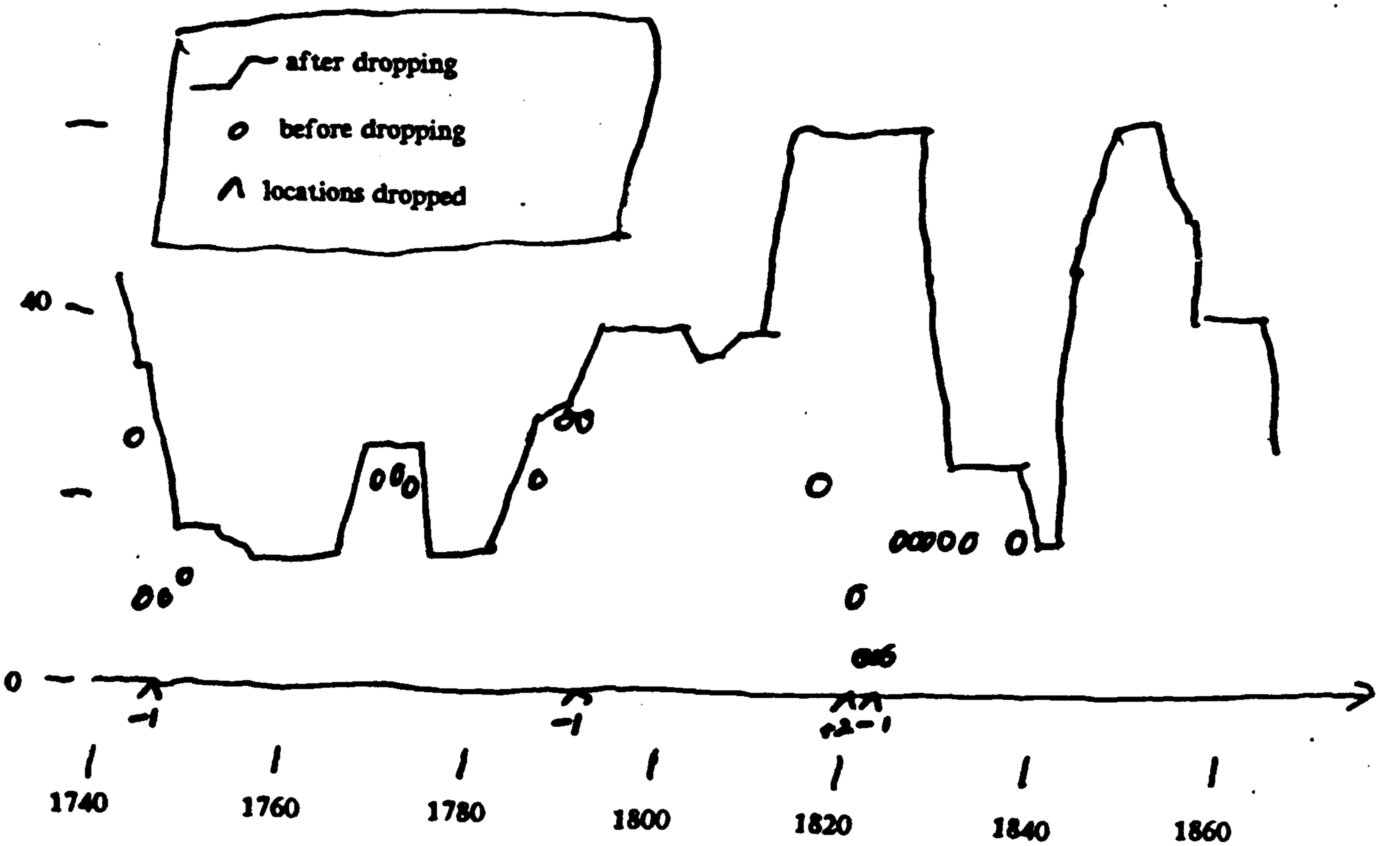
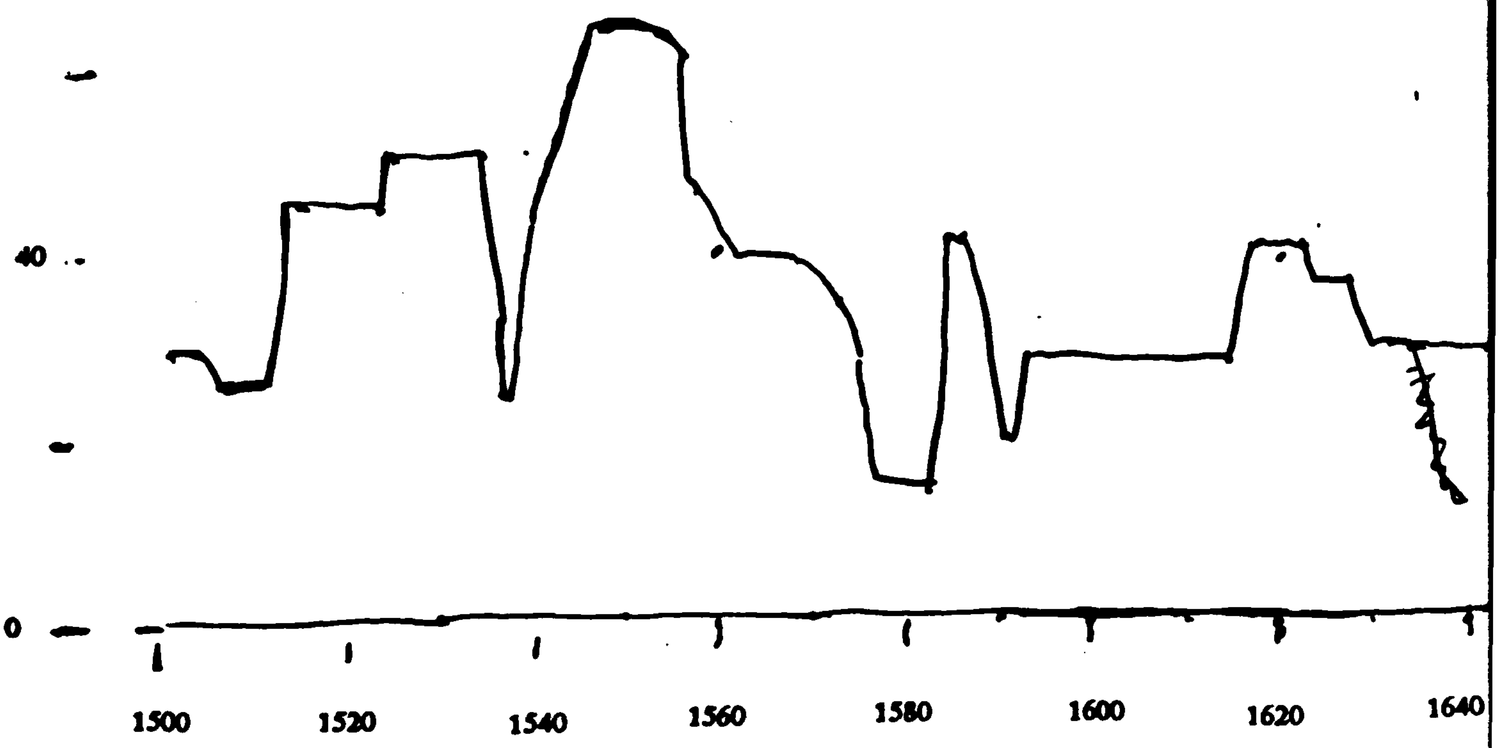


exhibit A13

Sample section of smooth of
peak-to-median-of-peaks

(compare with 1st section of exhibit A12)



We choose to apply first $D_{(3)}$ and then $D_{(1)}$, followed by "3" where

$$\Delta^2_{(3)}y_i = y_{i-3} - 2y_i + y_{i+3}$$
$$\Delta^2_{(1)}z_i = z_{i-1} - 2z_i + z_{i+1}$$
$$D_{(3)}y_i = y_i + \frac{1}{4}\Delta^2_{(3)}y_i, \quad \text{for most } i$$
$$= y_i, \quad \text{whenever} |\Delta^2_{(3)}y_i| \geq 3 \operatorname*{med}_j |\Delta^2_{(3)}y_j|$$
$$D_{(1)}z_i = z_i + \frac{1}{4}\Delta^2_{(1)}z_i, \quad \text{for most } i$$
$$= z_i \quad \text{whenever } |\Delta^2_{(1)}z_i| \geq 3 \operatorname*{med}_j |\Delta^2_{(1)}z_j|$$

The opening calculations are given in exhibit A14, and the points are plotted in exhibit A15. (The final "3" made very small changes in 3 places - - by interchanging two adjacent values - - in 1719-20, 1753-4 and 1757-8, in addition to the small displacement (at 1702) shown in exhibit A14.)

This final result is very smooth to the eye, except for 2 or 3 clean breaks (at 1736-7, 1784-5, and possibly 1718-9). It might well have been even smoother had we worked to one more decimal place. It shows the "Napoleonic hump" superimposed on a slowly rising trend (about 100 logarithmic units in 180 years, about 0.55% per year).

We *can* have visually very smooth results from simple, precisely defined smoothing techniques. Detrivializers can help a lot in this.

exhibit A14

Detrivialization of linear interpolates in XH3RP smooth.
(med $=|\Delta^2_{(3)}|$ from 1700 to 1866 is 2.5, med $=|\Delta^2_{(1)}|$ is 1.)

Year	XH3RP	(1)	$\Delta_{(3)}$	$\Delta^2_{(3)}$	$\frac{1}{4}$(same)	(2)	$\Delta_{(1)}$	$\Delta^2_{(1)}$	$\frac{1}{4}$(same)	(3)	3R
1690		92									
1		99									
2		106									
3	113		21	-21	*	113					
4		113	14	-14	*	113	0	-2	h		
95	113		7	-7	-2	111	-2	3	*		
6		113	0	-4	-1	112	1	0	0		
7		113	0	-8	*	113	1	-1	0		
8	113		0	-12	*	113	0	-6	*		
9		109	-4	-7h	-2	107	-6	4	*		
1700		105	-8	h	0	105	-2	-2	-h	104h	
1	101	-12	12	*	101	-4	h	0	101		
2	97h		-11h	11h	*	97h	-3h	5h	*	97h	99
3	97h		-7h	7h	2	99h	2	-2h	-h	99	
4		97h	0	6	1h	99	-h	-1	0	99	
05		97h	0	11h	*	97h	-1h	1h	h	98	
6	97h		0	17h	*	97h	0	1h	*	97h	
7		103h	6	5h	1h	105	7h	-4	*	105	
8		109	11h	-2h	-h	108h	3h	3	*	108h	
9	115		17h	-17h	*	115	6h	-6h	*	115	
1710		115	11h	-11h	*	115	0	0	0	115	

Notation: (1) interpolate between XH3RP points. $\Delta_{(3)} = y_i - y_{j-3}$.

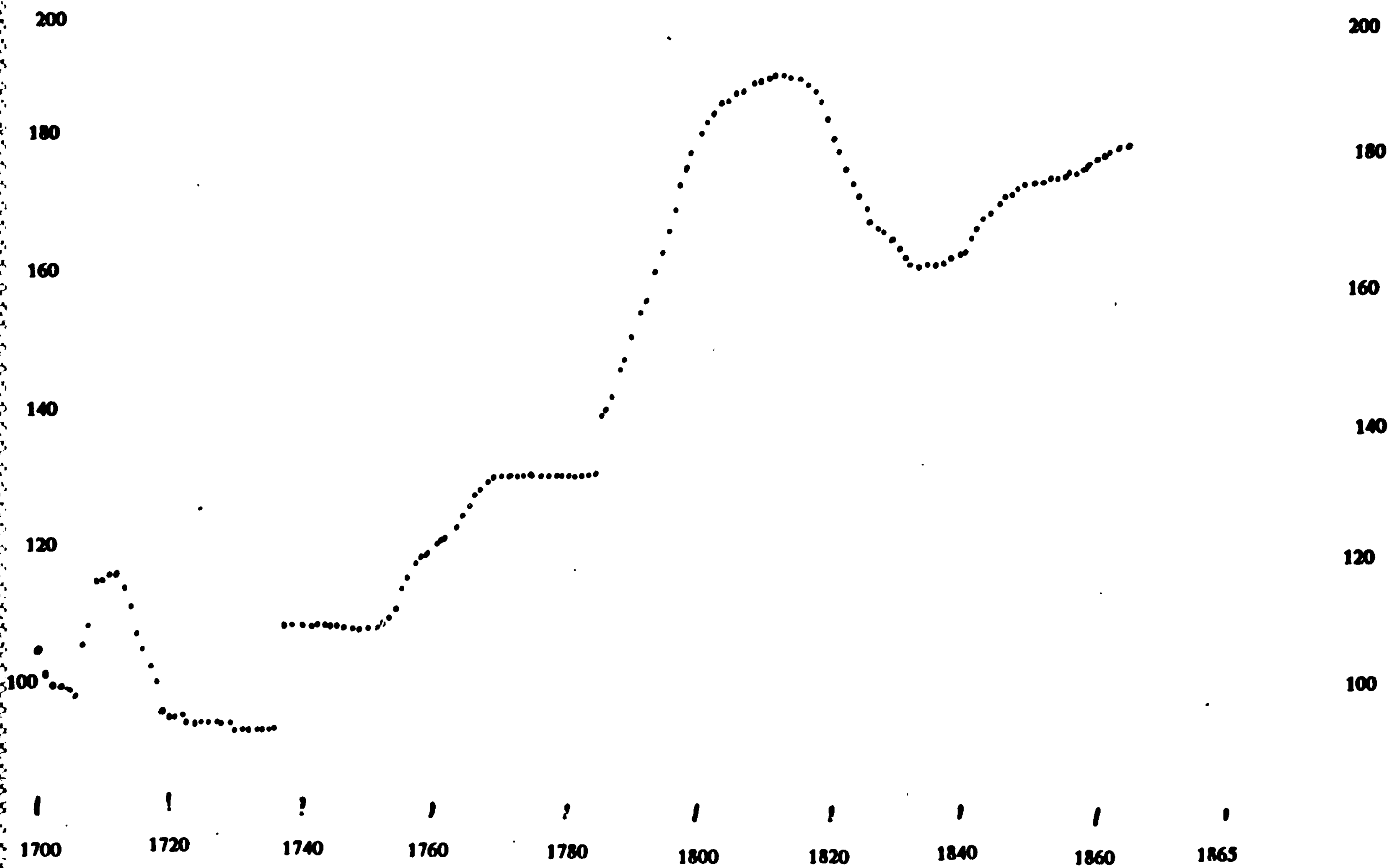
$\Delta^2_{(3)} = \Delta_{(3)}$ at i+3 MINUS Δ_3 at i; $\frac{1}{4}$(same) $= \frac{1}{4}\Delta^2_{(3)}$ except "*" (taken as zero,

when $|\Delta^2_{(3)}$ at $i| \geqslant 3$ median $|\Delta^2_{(3)}$ at $j|$ (2) $= D_{(3)}(1) =$ "(1)" plus "$\frac{1}{4}$(same)"

$\Delta_1 = z_i - z_{i-1}$, etc. (3) $= D_{(1)}(2)$.

exhibit A15

Detrivialized, modified XH3RP smooth of
the Beveridge wheat price series, 1700-1865



Appendix B

More on "local linear" smoothers

****** B1. Recent work at Stanford ******

The most recent work by Friedman and his collaborators involving local-linear fitting seems to be embodied in:

Jerome H. Friedman 1984. " A variable span smoother," *Technical Report No. 5*, November 1984, Laboratory for Computational Statistics, Department of Statistics, Stanford University.

John Alan McDonald and Art B. Owen 1984. "Smoothing with split linear fits," *Technical Report No. 7*, July 1984, Laboratory for Computational Statistics, Department of Statistics, Stanford University.

In Report No. 5, Friedman develops a locally-linear fit smoother using updating to make multiple choices of span, and eventually a variable span, computationally affordable. Absent exotic values, this smoother is reasonably attractive, both because of its performance against moderately difficult inputs and because the rationale for the various choices in its use are quite clearly explained. It is thus particularly important to emphasize that it is *neither* a robust *nor* a resistant smoother. (And that it does not take advantage of twicing.) All the local fits of straight lines are by least squares, and can be drawn far off by a relatively small number of exotic values.

A report dated 3 months earlier

Jerome W. Friedman, Gene H. Golub and Werner Stuetzle, "Project ORION, Final Report, August 1984 (ORION 026) Department of Statistics, Stanford University said (page 8, para 2) "In addition to the LCV smoother a rejection rule for outliers was developed. *If deemed necessary* (emphasis added), the LCV smoother can be preceded by application of the rejection rule to the data set, thus making the combined procedure resistant." It is far from clear *what* smoother Friedman *et al* *would recommend when.*

The smoother of #7 appears to be constructed to allow matching some of the properties of median-based smoothers — not indicating their abilities to deal with exotic values — within the framework of locally-linear least-square fitting. Its robustness is harder to assess than that of the previous smoother. By using a weighted combination of results for several windows, many of which extend only to the left or only to the right, it seems likely that this smoother has gained some robustness.

****** B2. Comments on "locally-linear" fitting ******

Discussions of "locally-linear" smoothing emphasize the geometric image of fitting local lines, but rarely come to the nub of the matter. As Friedman points out (1984, page 4), the simple moving average smoother has two serious shortcomings: "it does not reproduce straight lines if the abscissa values are not equispaced" and it has "bad behavior at the boundaries".

Why does the "locally-linear" smoother do better? Essentially because the fitted line is of the form

$$m_i + b_i(x-\bar{x}_i)$$

where m_i is the mean of the y's in the window associated with x_i, $\bar{x}_i$ that of the x's, and b_i is the corresponding slope. The value at x_i, which is the locally-linear smooth, is thus

$$m_i + b_i(x_i - \bar{x}_i)$$

where, away from the boundaries, $x_i - \bar{x}_i$ is often both quite small and an irregular function of i. The difference between "locally averaged" and "locally linear" smoothers is thus a correction term involving b_i as a multiplying factor. Thus it is appropriate to consider that all the complications involved in producing a well-tuned locally-linear smoother *at a fixed span* are concentrated in finding a reasonable sequence of estimates for a sequence of local slopes, which might be attacked in other ways. The remaining effort involves choice or mixing of spans, a matter of considerable importance.

****** B3. Cleveland's lowess ******

The basic reference still seems to be Cleveland 1979. Lowess, although (Cleveland 1979) discusses fitting polynomials of other degrees, uses *robust* locally-*linear* regression with *compound* weights -- products of robustness weights and window weights, the latter falling to zero at the furthest edge of the local window, which consists of the r points x-nearest to x_i, where $r = nf$ for some chosen $f < 1$.

(Cleveland, at page 834 (center right) worries about window-finding computations of order fn^2. Fortunately the division of the r points of a window into some on each side can be handled by bisection -- comparing $| x_i - x_A |$ and $|x_B - x_i|$ to learn which way to go, so that one window can be found in order $\log r = \log fn$ steps. After complete sorting, all windows can surely be found in order $n \log fn + \log n$ steps, which is order $n \log n$.

Cleveland further suggests (same paragraph) saving computation by grouping the x_i. It would seem as simple, and more effective, to group windows, grasping a window to minimize

$$\max\{ | x_i - x_A |, | x_B - x_{i+h} | \}$$

for h given and B-A = r + h, which can also be done by bisection. The single fit to this window can then be used for each of $x_i, x_{i+1}, \ldots, x_{i+h}$. All in all, the computational problems of lowess do not appear serious. (Other approaches seem to have been implemented.)

In using lowess it is important to realize that r = fn is for a tapered window not for a cookie-cutter window. Thus f =.5 in lowess is likely to correspond to something smaller, perhaps f = .3 for a Friedman smoother.

****** B4. Smelting ******

The estimation of local slopes, more precisely of their logarithms, is an essential of a procedure suggested by the author for allowing one quantity to guide the re-expression of another. This appears in J. W. Tukey 1981 "The use of smelting in guiding re-expression," *Modern Data Analysis*, A. F. Siegel and R. Launer, eds., Academic Press, New York, 83-102.

The basic approach involves, for an input of (u_i, v_i) pairs;

1) a fairly careful smoothing of the {u}, both by modification of values and by excision through replacement of successive i's with the same smoothed v by a single point (placed half-way between the extreme u's involved)

2) calculation of divided differences,

3) application of a median smoother to these divided differences (or, equivalently, to their logarithms) to identify which u-intervals should be combined (either because adjacent values are made equal or because adjacent values are interchanged)

Comment: the smoothed values obtained in (3) are *only* used to guide excision!

4) elimination (further excision) of the points whose removal will cause these intervals to be combined.

In the re-expression case, we want the signs of the divided differences to be constant, so we can work with the logarithms of their absolute values. And it is often reasonable to anticipate that the values underlying these logarithms will be monotone.

In the "slope for correcting moving means" application, however, we cannot be as sure of any of these conveniences. While stage (1) - - which uses vertical medians, 3RSS repeated to death, horizontal midextremes - - can probably be continued without much change (we might want to use horizontal means in the third subphase), we need to at least re-think the later phases.

This sort of approach might lead to an overall structure of the following form:

A) Smooth heavily, obtaining slope-estimates based upon excision and divided differences at a moderate number of places,

B) expand these results to all t by interpolation and extrapolation (linear?, constant?),

C) use the result as b's in adjusting moving average smoothers.

It is far from clear whether such an approach would prove to be an improvement.

Appendix C

A looming strategy

The example Appendix A and the discussion in Appendix B leaves us with an anticipation of one important place to go next. Given four things:

1) substantial amounts of data;

2) a desired to display the smooth to an eye (or eyes)

3) a belief that "lowess" or possibly a Friedman smoother would do moderately well, taking us quite a way to our goal, and

4) a recognition that it is no longer hard to do better (especially in terms of visual impression, perhaps even a little in terms of values read "off the curve")

We now find it natural to plan to follow, in order, the steps in the following multiphase strategy:

A) A robust initial fit, to strip off the most exotic values, replacing them by reasonable substitutes as an input to the next step.

B) A quality smoothometric fit, using "all the allowed principles of witchcraft" such as twicing, cross-validation and allowance for curvature.

C) Detrivialization or some other antirobust polish (may in part have been included in (B).)

Of these three phases, most of our attention needs to be directed toward (B), since we know a number of satisfactory ways to deal with (A), and expect (C) not to be difficult. Since we find it more convenient to discuss the issues in a more concrete context, we plan to discuss both the aspects needing modification and possible modifications, first for Friedman smoothers and then for Cleveland's lowess.

****** C1. Modifying Friedman's variable-span smoother. ******

This smoother (Friedman 1984, detailed reference in Appendix B) basically consists of three smoothers — woofer, midrange = middler, and tweeter — with smoothing of the qualitative results of cross-validation used to select a linear combination of adjacent smoothers. Exhibit C1 (Friedman's Figure 2b) shows the three smooths for an artificial example, whose points are tight to an oscillating curve at the left but loose to it on the right. Exhibit C2 (Friedman's Figure 2a) shows the resulting composite smooth.

As was to be expected, since the smooths are based on <u>un</u>twiced locally-<u>linear</u> <u>non</u>-robust fits, the woofer smooth fails to track hills and dales to any reasonable degree. It seems "a poor show" to use so unsatisfactory a smooth as competitor in the cross-validation. At least two natural cures are at hand.

*) We may twice (or maybe thrice) the woofer. [We can do this without increasing computing time by calculating the smooth at only every 3rd or 4th x-value, with the possible exception of x's near the boundaries. Since the woofer's span is n/2, we do not need closer detail, and can complete the calculation by linear interpolation.]

**) We may (a) fit a straight line, and (b) apply the woofer, then writing each observed value as

$$\text{observed} = (1 + K_i)(\text{woofer}) - K_i(\text{straight line})$$

with a different K_i for each data point we can smooth the values of K_i to obtain expansion factors, $\hat{K}_i$, and then a candidate smooth from

$$\text{smooth} = (1 + \hat{K}_i)(\text{woofer}) - \hat{K}_i(\text{straight line}).$$

(Limiting $|\hat{K}_i|$ to ≤ 2 will probably help.)

Either of these techniques should produce a reasonably improved candidate.

The middler (midrange) smooth does quite well in the example — although it

exhibit C1

Friedman's Figure 2b

Exhibit C1

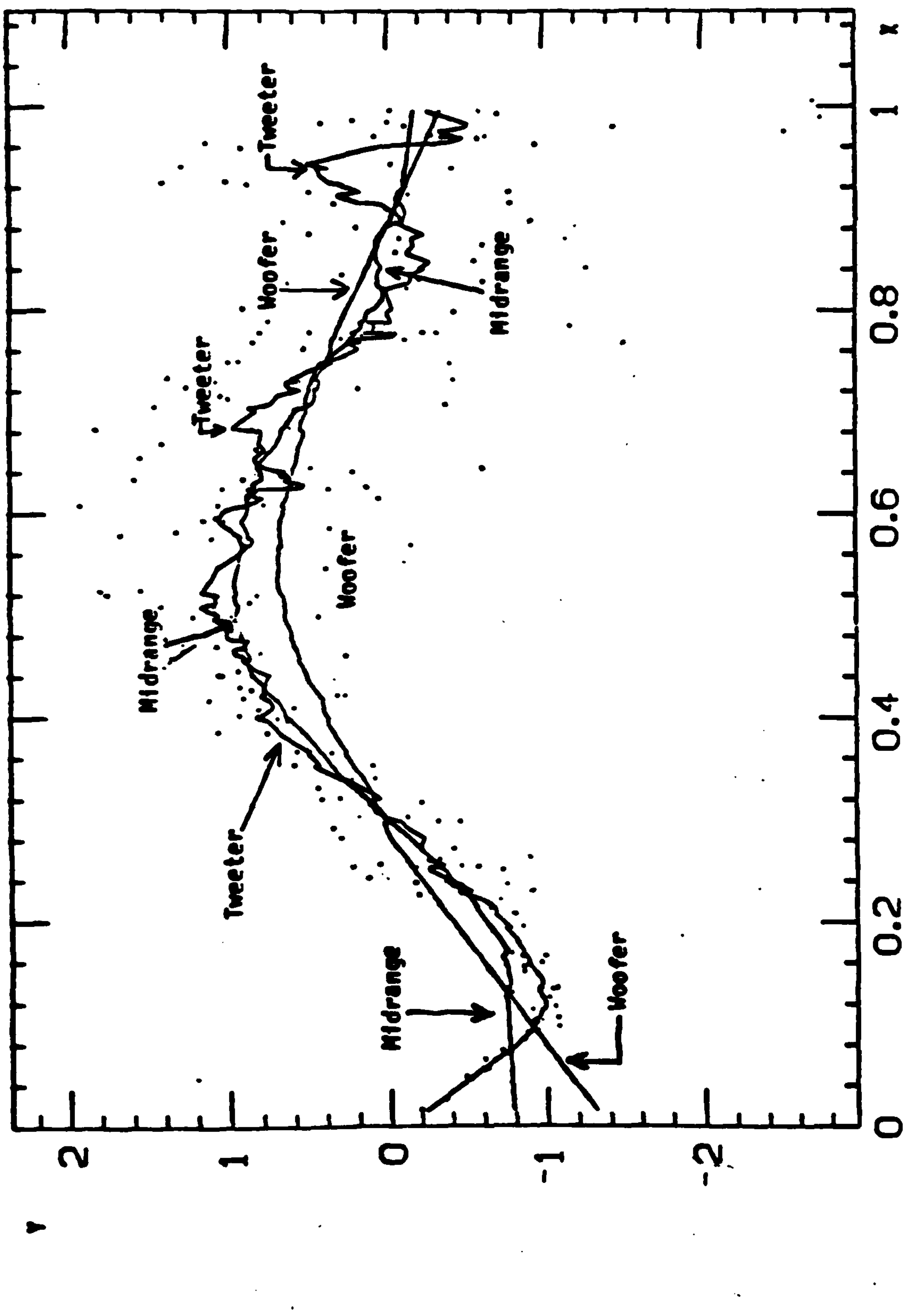


FIGURE 2b

exhibit C2

Friedman's Figure 2a

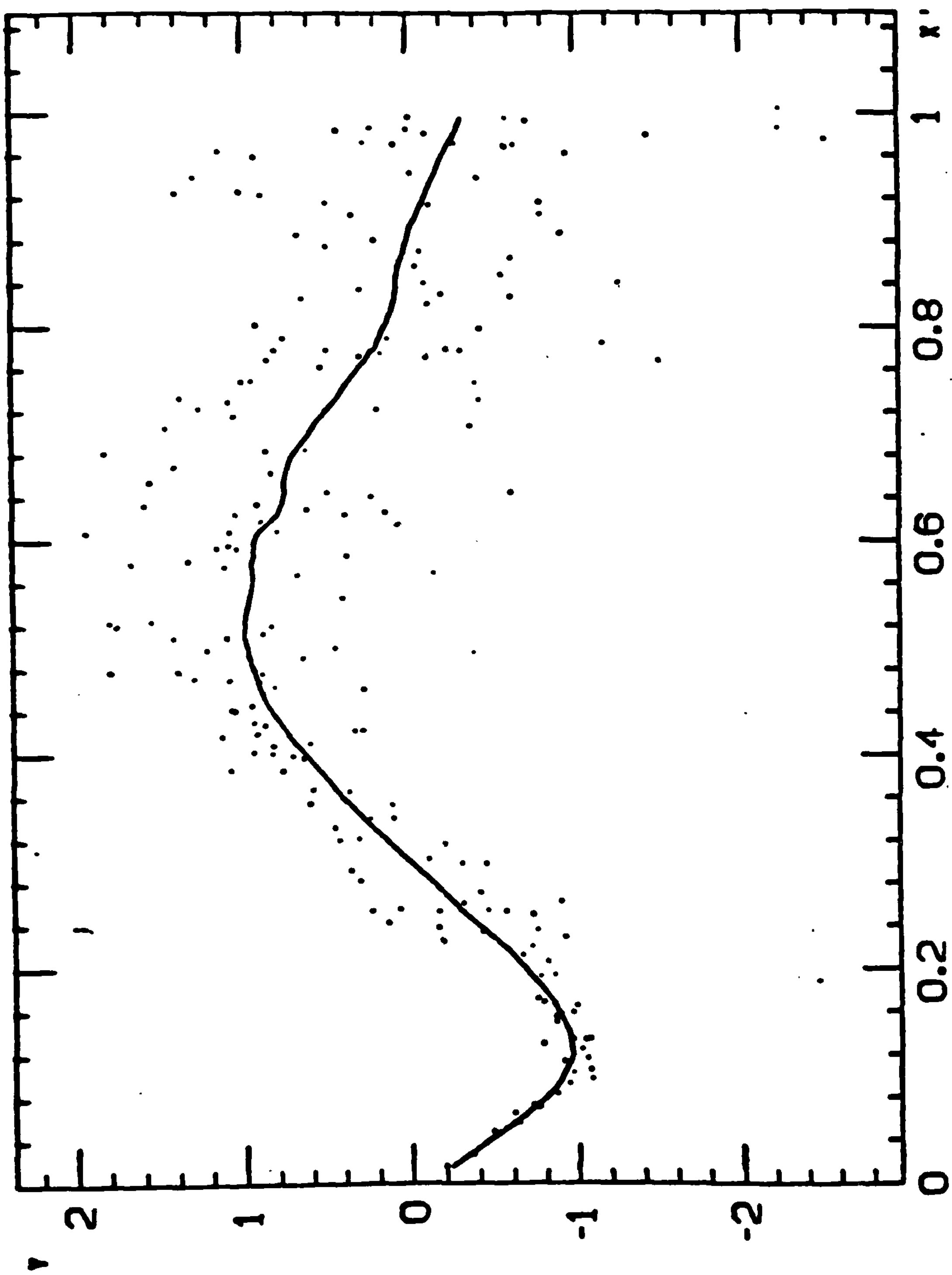


FIGURE 2a

seems unnecessarily rough. On the one hand, we might like to hope for a still better fit by (a) applying the middler to the rough from the modified woofer, and (b) taking the (modified) candidate smooth as

smooth by woofer PLUS smooth by middler of rough by woofer.

On the other, we might gain a little by detrivializing the candidate smooth (original or modified). Doing both could be a reasonable investment.

The tweeter smooth is mainly uncomfortable in terms of its irregularity. Detrivializing with $D_{(5)}$, $D_{(3)}$ and then $D_{(1)}$, in that order, should do no harm – and might well do good. Applying the tweeter to residuals from the middler smooth might also be desirable.

With 3 improved candidates, we can expect to do quite well by applying the Friedman technology of linearly combining candidates (his pp.8-9). It will probably be wise to smooth ($(|r_i(J)|)^{1/2)}$ rather than $|r_{(i)}(J)|$ against J, however. (Since we plan to get final visual smoothness by detrivialization, we ought not to have any need for a "bass (tone) control" (Friedman, pp. 9-10). We can thus avoid the difficulties shown in Friedman's figure 4b.)

****** C2. Curvature adjustment? ******

It may be that enough twicing was proposed in the last section to take care of the failure of "locally-linear" fits to allow for curvature. And it may not be that this is not so. Certainly the raw woofer is badly enough subject to curvature bias, that, if this is not fixed – for instance by either of the methods suggested in the last section – we should make some explicit allowance for curvature.

One way to do this is to:

1) find a high-grade visually-smooth smoothing $\{z_i\}$,

exhibit C3

Friedman's Figure 4b

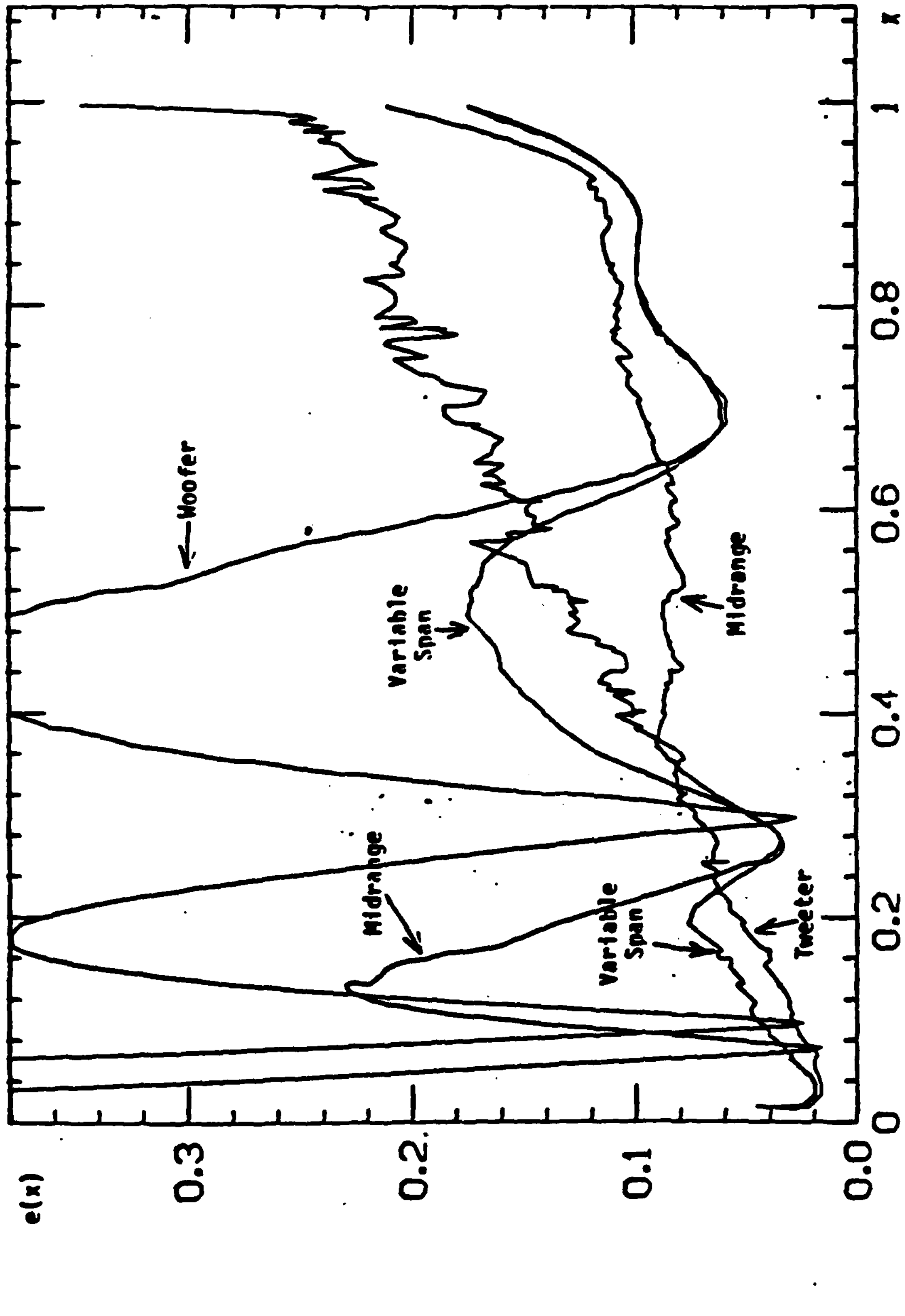


FIGURE 4b

2) reapply the whole smoothing process to $\{z_t\}$ obtaining $(Sz)_t$,

3) make a bias adjustment for the shift $(Sz)_t - z_t$, which means taking

$$z_t + (z_t - (Sz)_t) = 2z_t - (Sz)_t$$

as a bias-adjusted smooth.

While this last *step* may seem quite different from "twicing", a little algebra is illuminating: If z = Sy, then Sz = SSy, and 2z - Sz = 2Sy - SSy which approximates

$$S(2y - Sy) = S(y + Ry)$$

which approximates

$$Sy + SRy = \text{result of twicing.}$$

Both approximations would of course be exact equalities if S were superposable.

We do not yet have enough experience to know whether (or when) to prefer 2z - Sz to the result of twicing. (Even a selected convex linear combination of the two might be in order.)

In doing (2), it may be desirable to *force* the use of the same mixture of smooths, J(X) as was used in getting $\{z_t\}$.

***** C3. Improving Cleveland's lowess *****

As Cleveland's figures B and C clearly show:

1) Lowess is likely to benefit by further smoothing in the small (perhaps $D_{(5)}$ then $D_{(3)}$ then $D_{(1)}$ if the smooth is evaluated at 50-100 equispaced points).

2) We may want to limit the number of internal extremes in our smooth.

Re point 2, his figure C seems to have 9 such -- a smoothed-in-the-small version seems likely to retain 5 or 7 such -- for myself there are many instances (most smooths of circumstance response, for example) where I would like to limit the number of internal extremes to 0, $\leqslant 1$ or $\leqslant 2$ -- or, often, to each of these in turn.

(Time series smoothing or image smoothing would typically not call for such a limitation.)

Cleveland discusses, giving no detail for his algorithm, again on page 834, but on the lower left) the use of cross-validation to choose f. It would seem easy to modify the calculation to limit the number of internal extremes, after micropolishing, to 0, ≤ 1, or ≤ 2 (presumably available for f sufficiently close to 1). [The probable usefulness of such constrained cross-validation is no evidence against the possible existence of still better smooths subject to such constraints.]

At page 831 (lower right), Cleveland raises "the danger of inappropriate interpolation" when smoothed points are joined by straight lines. This is less of a worry than it might be, since Cleveland has just suggested calculating the fitted points at equal x-spacing. It can probably be changed from a loss to a gain by requiring connection if and only if, for the two adjacent points in question

$$|\text{slope}| \leq \text{med}\{\,|\text{slope}| \text{ all pairs of adjacent points}\}$$

(If two *adjacent segments are to be omitted the intermediate point should* be shown with a distinctive character.)

All in all, lowess should be reasonably satisfactory in its original form — and even more so modified. Its major disadvantages seem to be

a) roughness in the small, AND

b) no provision for limiting the number of internal extremes.

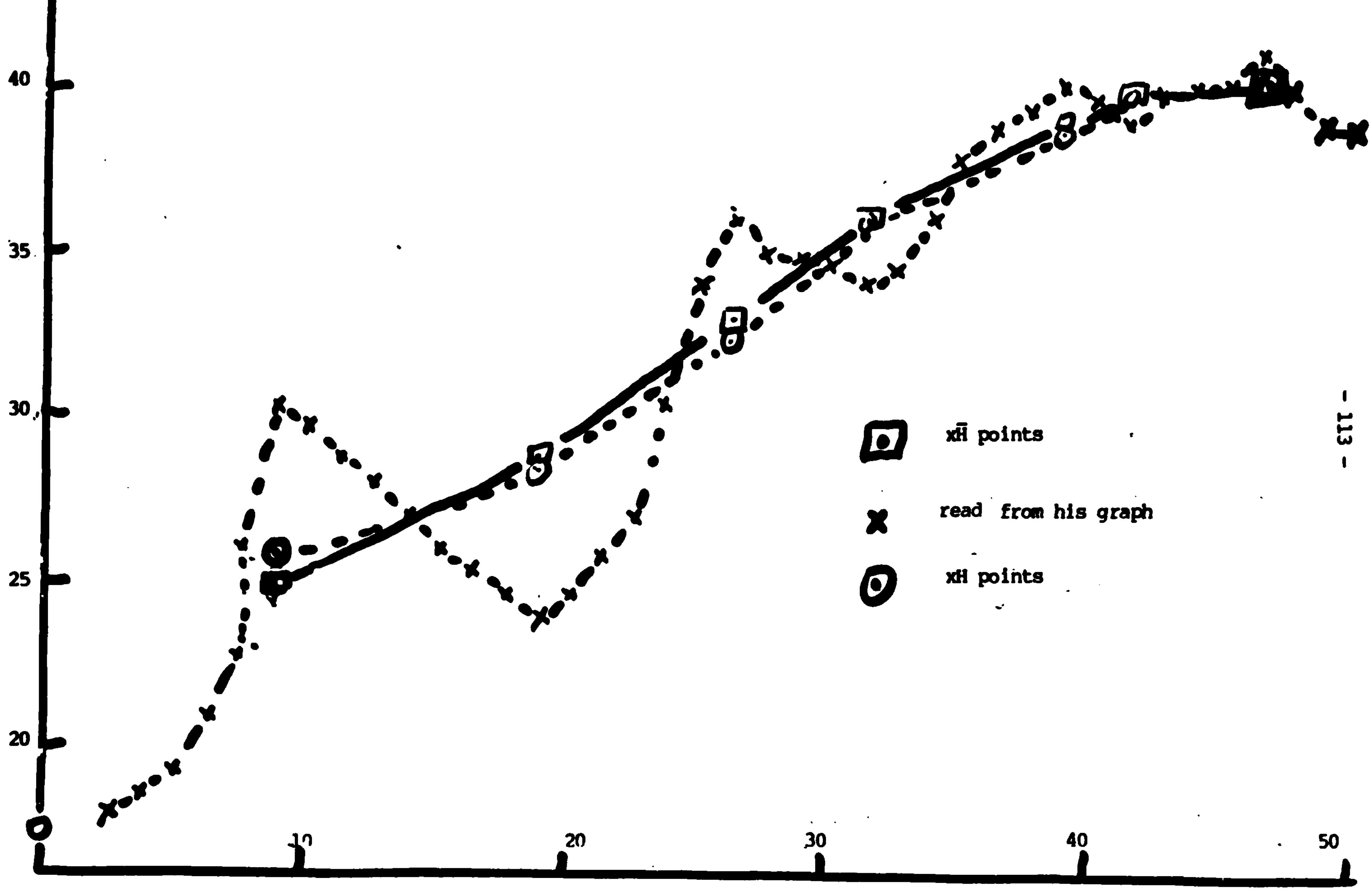
***** XH a possibility *****

When we look at Cleveland's Figure C, and remember the Beveridge series, we are tempted to try an XH calculation. Exhibit C4 shows:

1) points "read off the curve" for his figure C (symbol "x")

exhibit C4

xH and $x\bar{\bar{H}}$ smoothing of Cleveland' "figure C"



2) XH points, where H = (1/4, 1/2, 1/4) irrespective of spacing of extremes (symbol " · ")

3) $X\bar{\bar{H}}$, where $\bar{\bar{H}}$ averages one extreme with the linear interpolate of the adjacent extremes (symbol " + ")

4) various broken lines

It does seem that lowess with a small value of f may be usefully XH'd. (What to do near the ends is unclear.)

REFERENCES
(except for Appendices)

([ ] contain letters as per John W. Tukey's bibliographies in all volumes of The Collected Works)

Anderson, T. W. 1971. *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc., New York, pages 622-627.

Brillinger, D. R. 1970. "The identification of polynomial systems by means of higher order spectra," *J. Sound Vib. 12: 301-313.*

Cleveland, W. S. 1979. "Robust locally weighted regression and smoothing scatter plots," *J. Amer. Statist. Assoc., 74: 829-835.*

Cleveland, W. S. 1981. LOWESS: a program for smoothing scatterplots by robust locally weighted regression," *The American Statistician* 35:54

Gebski, V. and McNeil, D. 1984. "A refined method of robust smoothing," *J. Amer. Stat. Assoc.* Vol. 79, No. 387, 616-623.

Huang, T. S. 1981. *Two-dimensional signal processing II*, New York, Springer 1981.

Krystinik K. Bell, and Morgenthaler, S. 1981. "Comparison of the bioptimal curve with curves for two robust estimates," Technical Report No. 195, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey 08544, October 1981.

Mallows, C. L. 1980. "Some theory of non-linear smoothers," *Annals of Statistics* Vol. 8, 695-715.

Martin, R. D. and Thomson, D. 1982. *Proc. IEEE*, 70: 1097-1115.

Mazur, S., and Orlicz 1935. "Grundlegende Eigenschiaften der polynomischen operationen, Erste Mitteilungen," *Studia Mathematica 5:* 50-68 especially page 63.

Nodes, T. A. and Gallagher, N. C. 1982. "Median filters: some modifications and their properties," *IEEE Trans. Acoustics, Speech, and Signal Processing* ASSP. 30 739-746.

Schwartzschild, M. 1979. "New observation-outlier-resistant methods of spectrum estimation," Ph.D. dissertation, Department of Statistics, Princeton University, Princeton, New Jersey 08544.

Tukey, J. W. 1971[a]. *Exploratory Data Analysis* (Vol. 3 of Limited Preliminary Edition).

Tukey, J. W. (1974),1985[f]. "Nonlinear (nonsuperposable) methods for smoothing data," Chapter 22 in volume 2 of *The Collected Works of John W. Tukey* (ed. D. R. Brillinger) Wadsworth Publishing Company, Belmont, CA.

Tukey, J. W. 1977[a]. *Exploratory Data Analysis*, (First Edition, Addison-Wesley Publishing Company, Reading, Massachusetts, 688 pages.

Tukey, J. W. 1977[a]. *Exploratory Data Analysis*, (First Edition, Addison-Wesley Publishing Company, Reading, Massachusetts, 688 pages.

Tukey, J. W. 1979[g]. "Statistical Mapping: What should not be plotted," *Proceedings of the 1976 Workshop on Automated Catography*, DHEW Publication No. (PHS) 79-1254, 18-26.

Tukey, J. W. 1982[n]. "The use of smelting in guiding re-expression," *Modern Data Analysis*, (eds. A. F. Siegel and R. Launer), Academic Press, New York, 83-102.

Tukey, J. W. (1983)1984[j]. (14) "An introduction to the frequency analysis of time series, *The Collected Works of John W. Tukey: Volume 1, Time Series: 1949-1964*, (ed. D. R. Brillinger), Wadsworth Publishing Company, Belmont, CA, 503-650.

Tukey, J. W. and Tukey, P. A. 1981[e]. "Graphical display of data sets in 3 or more dimensions," Chapters 10, 11 and 12 *Interpreting Multivariate Data*, ed. V. Barnett, Chichester: John Wiley & Sons, Inc., New York, 189-275.

Velleman, P. 1975. "Robust non linear data smoothers - theory, definitions, and applications," Ph.D. dissertation, Department of Statistics, Princeton, New Jersey 08544. (See also 1975a, below)

Velleman, P. 1975a. "Robust non-linear data smoothing," Technical Report No. 89, (Series 2), Department of Statistics, Princeton University, Princeton, New Jersey 08544.

END

11-86

DTIC